

Winter 2012

Cross-correlation based performance measures for characterizing the influence of in-vehicle interfaces on driving and cognitive workload

Zeljko Medenica

University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

Recommended Citation

Medenica, Zeljko, "Cross-correlation based performance measures for characterizing the influence of in-vehicle interfaces on driving and cognitive workload" (2012). *Doctoral Dissertations*. 700.
<https://scholars.unh.edu/dissertation/700>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

**CROSS-CORRELATION BASED PERFORMANCE
MEASURES FOR CHARACTERIZING THE
INFLUENCE OF IN-VEHICLE INTERFACES ON
DRIVING AND COGNITIVE WORKLOAD**

BY

ŽELJKO MEDENICA

BS, Electrical and Computer Engineering, Faculty of Technical Sciences, Serbia, 2005

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

December, 2012

UMI Number: 3537822

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3537822

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.

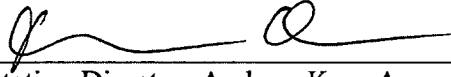
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

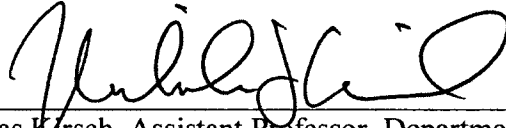
This dissertation has been examined and approved.



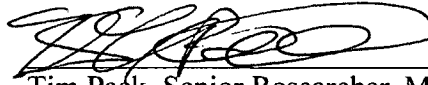
Dissertation Director, Andrew Kun, Associate
Professor, Department of Electrical and Computer
Engineering, University of New Hampshire



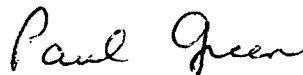
Thomas Miller, Professor, Department of Electrical
and Computer Engineering, University of New
Hampshire



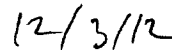
Nicholas Kirsch, Assistant Professor, Department of
Electrical and Computer Engineering, University of
New Hampshire



Tim Paek, Senior Researcher, Microsoft Research,
Electrical and Computer Engineering Department
Affiliate Professor



Paul Green, Research Professor, University of
Michigan, Transportation Research Institute



Date

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my advisor Dr. Andrew Kun for the support and guidance he showed me throughout the development of this dissertation.

I also would like to thank Dr. Thomas Miller, Dr. Nicholas Kirsch, Dr. Tim Paek and Dr. Paul Green, who have generously given their time and expertise for serving on my dissertation committee and provided me with valuable comments which improved this research.

I would like to thank all the faculty and students at Project54 for helping me with various aspects of this research.

Special thanks go to my beloved family, my mother, Nada, my father Čedomir, and my sister, Jelena, for their unequivocal support throughout my studies.

Above all, I would like to thank my loving wife Katarina for her constant support, care and great patience at all times. She helped me keep perspective on what is important in life and without her this achievement would not have been possible.

This work was supported by the US Department of Justice under grants 2006-DD-BX-K099, 2008-DN-BX-K221, 2009-D1-BX-K021 and 2010-DD-BX-K226, and by Microsoft Research.

TABLE OF CONTENTS

Acknowledgements.....	iii
Table of Contents.....	iv
List of tables.....	x
List of figures.....	xiv
List of acronyms	xxiv
Abstract.....	xxvi
Chapter 1 Introduction	1
1.1 Problem	5
1.1.1 Example Studies Reporting High Sensitivity of Average-based Measures	5
1.1.2 Example Studies Reporting Low Sensitivity of Average-based Measures	14
1.1.3 Problem Overview	25
1.2 Goals	29
1.3 Hypotheses	30
1.3.1 Hypothesis H1 – Quantifying Cumulative Secondary Task Engagements	34
1.3.2 Hypothesis H2 – Quantifying Instances of Secondary Task Engagements	39
1.3.3 Hypothesis H _{RP} – Ranking the Effects of Secondary Task Engagements.....	40
1.3.4 Hypothesis H3 – Analyzing Underlying Mechanisms	41
1.4 Testing Hypotheses	42

1.4.1 Testing H1 and H2 – Cumulative and Instance-based Quantifications of Secondary Task Engagements	42
1.4.2 Testing H _{RP} – Ranking the Effects of Secondary Task Engagements	44
1.4.3 Testing H3 – Analyzing Underlying Mechanisms	45
1.5 Dissertation Organization.....	46
Chapter 2 Background	47
2.1 Driver Distractions	48
2.2 Cognitive Load.....	53
2.2.1 Performance-based Measures	63
2.2.2 Physiological Measures	67
2.2.3 Subjective Measures	72
2.2.4 Criteria for the Selection of Workload Measures	76
2.3 Experimental Method.....	80
2.3.1 Experimental Apparatus	80
2.3.2 Experimental Approach.....	83
2.4 Studies Employing Cross-Correlation Function	85
2.5 Studies Employing Regression Analysis	88
Chapter 3 Cross-correlation Method.....	91
3.1 Hypotheses Addressed in this Chapter.....	91
3.1.1 General Terminology.....	92
3.1.2 Requirements of the Method	97
3.1.3 Definition of the Method	106
3.1.4 Algorithm	113

3.1.5 Ranking Cross-Correlation Results	118
3.2 Studies Implementing Cross-Correlation Method	122
3.2.1 Exploring Augmented Reality Navigation Aids.....	123
3.2.2 Highway Driving and iPod Interactions.....	147
3.2.3 General Discussion of the Results.....	168
Chapter 4 Mechanisms Underlying Cross-correlation Results.....	180
4.1 City Driving and iPod Interactions	181
4.2 Observing Effects of Driving Environment through Reference Studies.....	200
4.2.1 Effects of Driving Environment on Visual Attention.....	201
4.2.2 Effects of Driving Environment on Average-Based Driving Performance Measures.....	205
4.2.3 Effects of Driving Environment on Cross-Correlation Results.....	209
4.2.4 Comparing Average-Based Driving Performance Measures and Cumulative Cross-Correlation Results.....	213
4.3 Obtaining Predictors of Cross-Correlation Results.....	214
4.3.1 Describing Visual Attention.....	215
4.3.2 Describing Driving Performance.....	217
4.3.3 Data Collection.....	219
4.3.4 Data Conditioning for Regression Analysis	223
4.3.5 Creating Regression Models for Reference Studies.....	233
4.3.6 Modeling the Effect of the Driving Environment	253
4.3.7 General Discussion and Future Direction.....	266
Chapter 5 Conclusions and Future Work.....	279

5.1 Primary Contributions	281
5.1.1 Addressing Goal 1	282
5.1.2 Addressing Goal 2	286
5.1.3 Addressing Goal 3	288
5.2 Secondary Contributions	291
5.3 Extensions of the Current Work	292
5.4 Applicability of the Current Work	295
APPENDIX A Data Synchronization	302
A.1 Hardware Setup	304
A.2 Software Setup	306
A.2.1 Configuration Files	308
A.2.2 Log Files	309
A.2.3 Establishing Communication between SymConnect and Driving Simulator	309
A.2.4 Main Window Options	311
A.2.5 Settings Menu Options	315
A.3 Data Synchronization	319
A.4 Simple Experiment Example	325
A.5 Synchronizing Audio Recordings	329
APPENDIX B Experimental Apparatus	334
B.1 Driving Simulator	334
B.2 Eye-tracker	336
B.3 Physiological Monitor	339

B.4 Institutional Review Board Form	341
B.5 NASA-TLX Description Presented to Participants	347
APPENDIX C Checking Assumptions behind Cross-Correlation Models	348
C.1 Testing Cumulative Steering Wheel Angle Cross-Correlation Models for Highway Study	350
C.2 Testing Cumulative Lane Position Cross-Correlation Models for Highway Study	351
C.3 Testing Cumulative Steering Wheel Angle Cross-Correlation Models for City Study	352
C.4 Testing Cumulative Lane Position Cross-Correlation Models for City Study	353
C.5 Testing Per-Glance Steering Wheel Angle Cross-Correlation Models for Highway Study	354
C.6 Testing Per-Glance Lane Position Cross-Correlation Models for Highway Study	355
C.7 Testing Per-Glance Steering Wheel Angle Cross-Correlation Models for City Study	356
C.8 Testing Per-Glance Lane Position Cross-Correlation Models for City Study	357
C.9 Testing Cumulative Steering Wheel Angle Cross-Correlation Models for Pooled Highway and City Studies	358
C.10 Testing Cumulative Lane Position Cross-Correlation Models for Pooled Highway and City Studies	359
C.11 Testing Per-Glance Steering Wheel Angle Cross-Correlation Models for Pooled Highway and City Studies	360

C.12 Testing Per-Glance Lane Position Cross-Correlation Models for Pooled Highway and City Studies	361
References	362

LIST OF TABLES

Table 2.1 Interaction types explored in four preliminary studies.	51
Table 2.2 Resource allocation for three example studies.	56
Table 2.3 Average-based driving performance measures employed in studies presented in this dissertation.....	67
Table 2.4 Physiological measures employed in studies presented in this dissertation.	71
Table 2.5 Subjective measures used in studies presented in this dissertation.	76
Table 2.6 Driving types employed in studies presented in this dissertation.	84
Table 3.1 PDT on the road, LCD and the rest of the cabin as a function of PND type...130	
Table 3.2 Level of agreement with two preferential statements.	133
Table 3.3 Statistical comparisons between cumulative cross-correlation results for three PNDs.	140
Table 3.4 Statistical comparisons between per-glance cross-correlation results for three PNDs.	146
Table 3.5 Results of statistical analyses for all dependent driving variables.....	156
Table 3.6 Statistical comparisons between cumulative cross-correlation results for three interaction types.	161
Table 3.7 Statistical comparisons between per-glance cross-correlation functions for three interaction types.	167

Table 4.1 Statistical analyses of average-based driving performance measures.	188
Table 4.2 Results of statistical comparisons between cumulative cross-correlation functions for B, E and D conditions.....	192
Table 4.3 Results of statistical comparisons between per-glance cross-correlation functions for B, E and D conditions.....	197
Table 4.4 Results of two-way ANOVAs for cumulative and per-glance cross-correlation results.	210
Table 4.5 Comparing magnitudes of most prominent cross-correlation peaks for two reference studies.....	211
Table 4.6 Relative differences between interface types in both environments.....	212
Table 4.7 Comparison of significant effects detected using average-based measures for the two reference studies.....	214
Table 4.8 Overview of the proposed independent variables and their corresponding transformations used in regression analyses.	222
Table 4.9 Exponents of power transformations used for normalizing the data.	225
Table 4.10 Highway study: Coefficient estimates for CML_M1_SWA model.	238
Table 4.11 Highway study: Coefficient estimates for CML_M2_SWA model.	239
Table 4.12 Highway study: Coefficient estimates for CML_M1_LP model.....	239
Table 4.13 Highway study: Coefficient estimates for CML_M2_LP model.....	240
Table 4.14 City study: Coefficient estimates for CML_M1_SWA model.	241
Table 4.15 City study: Coefficient estimates for CML_M2_SWA model.	242
Table 4.16 City study: Coefficient estimates for CML_M1_LP model.	242
Table 4.17 City study: Coefficient estimates for CML_M2_LP model.	243

Table 4.18 Highway study: Coefficient estimates for PG_M_SWA model.....	244
Table 4.19 Highway study: Coefficient estimates for PG_M_LP model.	245
Table 4.20 City study: Coefficient estimates for PG_M_SWA model.....	246
Table 4.21 City study: Coefficient estimates for PG_M_LP model.....	247
Table 4.22 Standardized beta coefficients for individual reference studies for model CML_M1_X.....	249
Table 4.23 Standardized beta coefficients for individual reference studies for model CML_M2_X.....	249
Table 4.24 Standardized beta coefficients for individual reference studies for model PG_M_X.	251
Table 4.25 Exponents of power transformations used for normalizing the data for pooled reference studies.....	255
Table 4.26 Coefficient estimates for CML_M1_SWA model.....	259
Table 4.27 Coefficient estimates for CML_M2_SWA model.....	260
Table 4.28 Coefficient estimates for CML_M1_LP model.	261
Table 4.29 Coefficient estimates for CML_M2_LP model.	261
Table 4.30 Coefficient estimates for PG_M_SWA model.	262
Table 4.31 Coefficient estimates for PG_M_LP model.....	263
Table 4.32 Standardized beta coefficients for pooled reference studies for model CML_M1_X.....	264
Table 4.33 Standardized beta coefficients for pooled reference studies for model CML_M2_X.....	264

Table 4.34 Standardized beta coefficients for pooled reference studies for model PG_M_X.	264
Table 4.35 Observed cumulative cross-correlation results in the navigation study.	269
Table 4.36 Predicted cumulative cross-correlation results using CML_M1_X model. ..	270
Table 4.37 Comparison of observed and normalized predicted cumulative cross- correlation results.	271
Table 4.38 Observed per-glance cross-correlation results in the navigation study.	272
Table 4.39 Predicted per-glance cross-correlation results using PG_M_X model.	272
Table 4.40 Observed and predicted cross-correlation results for both reference studies. Predicted results for each study were obtained using models from the opposite study.	275
Table 5.1 Coefficients of determination obtained between NASA-TLX and cumulative cross-correlation results for three studies.	284
Table 5.2 Coefficients of determination obtained between average skin conductance and cumulative cross-correlation results.	285
Table B.3 Description of NASA-TLX scales.	347

LIST OF FIGURES

Figure 1.1 Participant manually adjusting channels on the radio inside the simulator.....	8
Figure 1.2 Box plots of variances for lane position, steering wheel angle and velocity.	9
Figure 1.3 Comparison of lane position for manual (left) and speech interaction (right) for one example participant.	10
Figure 1.4 Mean NASA-TLX workload score (error bars represent ± 1 SD).	11
Figure 1.5 Mean steering wheel angle (left) and lane position (right) variances.	13
Figure 1.6 7” LCD screen simulating a map-based standard PND.....	16
Figure 1.7 Experimental setup inside the simulator cabin.	17
Figure 1.8 Mean variances of lane position (left) and steering wheel angle (right).	19
Figure 1.9 PDT on the outside world.....	19
Figure 1.10 Simulated two-lane city road with unexpected event.....	20
Figure 1.11 Changes in PDT on the outside world (left) and SPND (right) based on distances from intersections.	22
Figure 1.12 Problem illustration using a specific example.....	26
Figure 1.13 Cross-correlation results comparing standard and spoken-only PND as obtained using initial cross-correlation method.	36
Figure 2.1 Inverted U-shape relationship between performance and workload (arousal).	53
Figure 3.1 Model of our driving simulator’s cabin employed in our eye-tracker.....	95

Figure 3.2 Pictorial explanation of the EGS sequence.	96
Figure 3.3 View of the participant from the camera mounted on the dashboard.....	98
Figure 3.4 Illustration of the glance filtering procedure.	101
Figure 3.5 Converting glances from the 60 Hz time scale (eye-tracker) to the 10 Hz time scale (driving simulator).	104
Figure 3.6 Searching for the closest sample on the 10 Hz time scale.....	105
Figure 3.7 Differences in rising and falling edges between two time scales.....	106
Figure 3.8 Example cumulative and per-glance cross-correlation results using two simulated sequences.	111
Figure 3.9 Pseudo-code for estimating cross-correlation results.	115
Figure 3.10 Pseudo-code for calculating statistical significance.	118
Figure 3.11 Pseudo-code for testing statistically significant differences between experimental conditions.	122
Figure 3.12 AR navigation aid shown from driver's perspective.	123
Figure 3.13 SV navigation aid displayed on LCD.	124
Figure 3.14 SPND navigation aid displayed on LCD.	125
Figure 3.15 Experimental setup inside the simulator cabin.	125
Figure 3.16 Simulated two-lane city road with a pedestrian emerging from behind a parked vehicle.	127
Figure 3.17 Simulated route with the segments selected for analysis.	128
Figure 3.18 Average duration (left) and number (right) of glances directed off road for the three PNDs per segment.	132

Figure 3.19 Average variances of lane position (upper left), steering wheel angle (upper right) and (bottom left) and average velocity (bottom right).	135
Figure 3.20 Cumulative steering wheel angle cross-correlation functions calculated for AR, SV and SPND.	136
Figure 3.21 Cumulative lane position cross-correlation functions calculated for AR, SV and SPND.	137
Figure 3.22 Magnitudes of prominent peaks in cumulative R_{sw} (upper graph) and R_{lp} (lower graph) vs. NASA-TLX score for AR, SV and SPND.	142
Figure 3.23 Per-glance steering wheel angle cross-correlation functions calculated for AR, SV and SPND.	143
Figure 3.24 Per-glance lane position cross-correlation functions calculated for AR, SV and SPND.	144
Figure 3.25 Example raw data for lane position and steering wheel angle.	145
Figure 3.26 Amplitude spectra for example lane position and steering wheel angle data.	145
Figure 3.27 Experimental setup inside the simulator cabin.	148
Figure 3.28 iPod interactions participants performed during the experiment.	150
Figure 3.29 Average PDT on the forward road.	153
Figure 3.30 Average duration (left) and number (right) of glances directed off road.	154
Figure 3.31 Average NASA-TLX score.	155
Figure 3.32 Average variances of lane position (upper left), steering wheel angle (upper right) and velocity (bottom left) and average velocity (bottom right).	155

Figure 3.33 Cumulative steering wheel angle cross-correlation functions calculated for D, B and E conditions.	158
Figure 3.34 Cumulative lane position cross-correlation functions calculated for D, B and E conditions.....	159
Figure 3.35 Magnitudes of prominent peaks in cumulative R_{sw} (upper graph) and R_{lp} (lower graph) vs. NASA-TLX score for B, E and D conditions.....	163
Figure 3.36 Per-glance steering wheel angle cross-correlation functions calculated for D, B and E conditions.	164
Figure 3.37 Per-glance lane position cross-correlation functions calculated for D, B and E conditions.	164
Figure 3.38 Example raw data for lane position and steering wheel angle.	166
Figure 3.39 Amplitude spectra for example lane position and steering wheel angle data.	166
Figure 4.1 Simulated city environment.....	182
Figure 4.2 Average PDT on the forward road.	184
Figure 4.3 Average duration (left) and number (right) of glances directed off-road.....	185
Figure 4.4 Average NASA-TLX score.	186
Figure 4.5 Average physiological measures: skin conductance (left) and heart rate (right).....	186
Figure 4.6 Average variances of lane position (upper left), steering wheel angle (upper right), velocity (lower left) and average velocity (lower right).....	188
Figure 4.7 Cumulative steering wheel angle cross-correlation functions calculated for B, E and D conditions.	190

Figure 4.8 Cumulative lane position cross-correlation functions calculated for B, E and D conditions.	190
Figure 4.9 Magnitudes of the most prominent peaks of cumulative cross-correlation functions <i>Rsw</i> and <i>Rlp</i> vs. NASA-TLX score for B, E and D conditions.	193
Figure 4.10 Magnitudes of the most prominent peaks of cumulative cross-correlation functions <i>Rsw</i> and <i>Rlp</i> vs. average skin conductance for B, E and D conditions.	194
Figure 4.11 Per-glance steering wheel angle cross-correlation functions calculated for B, E and D conditions.	195
Figure 4.12 Per-glance lane position cross-correlation functions calculated for B, E and D conditions.	196
Figure 4.13 Average PDT on the forward road observed in city and highway driving while interacting with the iPod.	202
Figure 4.14 Average glance duration directed off-road in city and highway driving while interacting with the iPod.	203
Figure 4.15 Average number of glances directed off-road in city and highway driving while interacting with the iPod.	204
Figure 4.16 Average steering wheel angle variance in city and highway driving while interacting with the iPod.	206
Figure 4.17 Average lane position variance in city and highway driving while interacting with the iPod.	206
Figure 4.18 Average velocity variance in city and highway driving while interacting with the iPod.	207

Figure 4.19 Illustration of the relationships between steering wheel angle, lane position and vehicle heading	218
Figure 4.20 Highway study: distributions of independent variables used for modeling cumulative cross-correlation results	227
Figure 4.21 Highway study: distributions of dependent variables used for modeling cumulative cross-correlation results	228
Figure 4.22 Highway study: distributions of independent variables used for modeling per-glance cross-correlation results.	229
Figure 4.23 Highway study: distributions of dependent variables used for modeling per-glance cross-correlation results.	229
Figure 4.24 City study: distributions of independent variables used for modeling cumulative cross-correlation results	230
Figure 4.25 City study: distributions of dependent variables used for modeling cumulative cross-correlation results	231
Figure 4.26 City study: distributions of independent variables used for modeling per-glance cross-correlation results.	231
Figure 4.27 City study: distributions of dependent variables used for modeling per-glance cross-correlation results.	232
Figure 4.28 Distributions of dependent and independent variables used for modeling per-glance cross-correlation results for pooled highway and city studies	256
Figure 4.29 Distributions of independent variables used for modeling cumulative cross-correlation results for pooled highway and city studies	257

Figure 4.30 Distributions of dependent variables used for modeling cumulative cross-correlation results for pooled highway and city studies.....	258
Figure 5.1 Driver (left) and other conversant (right).	297
Figure 5.2 Cross-correlation functions for all six drivers (left) and for the four drivers whose results clearly supported our hypothesis (right).....	298
Figure A.3 Synchronizing experimental equipment.	303
Figure A.4 Synchronization box.	304
Figure A.5 Inside view of the synchronization box.	305
Figure A.6 Schematic of the synchronization box.	306
Figure A.7 SymConnect's main window.	307
Figure A.8 "Port and IP" dialog.	308
Figure A.9 "Settings" menu.	308
Figure A.10 SymConnect window while sending commands to the simulator.	311
Figure A.11 Sending a message to Project54's speechio application.	315
Figure A.12 Send commands window.	316
Figure A.13 Contents of the file "user_initiative.txt".	316
Figure A.14 COM port selection.	320
Figure A.15 Synchronization is enabled by activating the COM port.	321
Figure A.16 Waiting for sync signal (left) and signal received (right).	323
Figure A.17 Sample main SymConnect's "measurements.txt" file.	328
Figure A.18 Sample secondary SymConnect's "measurements.txt" file.	329
Figure A.19 Dashboard light signal.	331
Figure A.20 Breadboard with the astable multivibrator circuit.	332

Figure A.21 Audio synchronization circuit schematic.	333
Figure B.22 High fidelity driving simulator.	335
Figure B.23 Eye-tracker mounted on top of the dashboard.	336
Figure B.24 A view of the participant as seen by the eye-tracker.	337
Figure B.25 Fitted pupil diameter.	338
Figure B.26 Virtual model of the car cabin.	338
Figure B.27 Physiological monitor and the corresponding sensors for skin conductance and heart rate.	339
Figure B.28 Skin conductance electrodes embedded in a glove and electrode straps.	340
Figure C.29 Highway study: Distributions of studentized residuals for CML_M1_SWA (left) and CML_M2_SWA (right).	350
Figure C.30 Highway study: Residuals versus predicted plots for CML_M1_SWA (left) and CML_M2_SWA (right).	350
Figure C.31 Highway study: Distributions of studentized residuals for CML_M1_LP (left) and CML_M2_LP (right).	351
Figure C.32 Highway study: Residuals versus predicted plots for CML_M1_LP (left) and CML_M2_LP (right).	351
Figure C.33 City study: Distributions of studentized residuals for CML_M1_SWA (left) and CML_M2_SWA (right).	352
Figure C.34 City study: Residuals versus predicted plots for CML_M1_SWA (left) and CML_M2_SWA (right).	352
Figure C.35 City study: Distributions of studentized residuals for CML_M1_LP (left) and CML_M2_LP (right).	353

Figure C.36 City study: Residuals versus predicted plots for CML_M1_LP (left) and CML_M2_LP (right).....	353
Figure C.37 Highway study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_SWA model.	354
Figure C.38 Highway study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_LP model.	355
Figure C.39 City study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_SWA model.	356
Figure C.40 City study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_LP model.	357
Figure C.41 Distributions of studentized residuals for CML_M1_SWA (left) and CML_M2_SWA (right) models.	358
Figure C.42 Residuals versus predicted plots for CML_M1_SWA (left) and CML_M2_SWA (right) models.	358
Figure C.43 Distributions of studentized residuals for CML_M1_LP (left) and CML_M2_LP (right) models.	359
Figure C.44 Residuals versus predicted plots for CML_M1_LP (left) and CML_M2_LP (right).....	359
Figure C.45 Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_SWA model.	360
Figure C.46 Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_LP model.....	361

LIST OF ACRONYMS

AAVC	Average absolute value of change
ANOVA	Analysis of variance
AR	Augmented reality
AVC	Absolute value of change
CML	Cumulative
DPS	Driving performance sequence
EDR	Electro-dermal response
EGS	Eye-glance sequence
GD	Glance duration
GUI	Graphical user interface
HDD	Head-down display
HR	Heart rate
HUD	Head-up display
IVIS	In-vehicle information system

LAN	Local area network
LCD	Liquid crystal display
LP	Lane position
M	Mean (average) value
MP3	MPEG-1 Audio layer 3
NG	Number of glances
PDT	Percent dwell time
PDT_AFR	Percent dwell time away from road
PG	Per-glance
PND	Personal navigation device
PTT	Press-to-talk button
SC	Skin conductance
SD	Standard deviation
SPND	Standard map-based personal navigation device
SUI	Speech user interface
SV	Street view
SWA	Steering wheel angle
VH	Vehicle heading

ABSTRACT

CROSS-CORRELATION BASED PERFORMANCE MEASURES FOR CHARACTERIZING THE INFLUENCE OF IN-VEHICLE INTERFACES ON DRIVING AND COGNITIVE WORKLOAD

by

ŽELJKO MEDENICA

University of New Hampshire, December, 2012

Driving is a cognitively loading task which requires drivers' full attention and coordination of both mind and body. However, drivers often engage in side activities which can negatively impact safety. A typical approach for analyzing the influences of side activities on driving is to conduct experiments in which various driving performance measures are collected, such as steering wheel angle and lane position. Those measures are then transformed, typically using means and variances, before being analyzed statistically. However, the problem is that those transformations perform averaging of the

acquired data, which can result in missing short, but important events (such as glances directed off-road). As a consequence, statistically significant differences may not be observed between the tested conditions. Nevertheless, just because the influences of in-vehicle interactions do not show in the averages, it does not mean that they do not exist or should be neglected, especially if the nature of the interactions is such that they can be performed frequently (for example, with an infotainment system). This can create a false conclusion about the lack of influence of the tested side activity on driving.

The main contribution of this research is in developing two new performance measures inspired by the mathematical function of cross-correlation: one which evaluates the cumulative effect and the other which evaluates the effects of individual instances of in-vehicle interactions on driving and cognitive load. The results from three driving simulator studies demonstrate that our cumulative measure provides more sensitivity to the effects of in-vehicle interactions, even when they are not detected through average-based measures. Additionally, our instance-based measure provides a low-level insight into the nature of the influence of individual in-vehicle interactions. Both measures produce results that can be ranked, which allows determining the relative size of the effect that various in-vehicle interactions have on driving. Finally, we demonstrate a set of variables which can be used for predicting the cumulative and instance-based results. This predictive ability is important, because it may allow obtaining quick simulation results without performing actual experiments, which can be used in the early stages of an interface or experiment design process.

CHAPTER 1

INTRODUCTION

In recent years we have seen a major increase in research concerned with driver distraction and the influence of various in-vehicle devices on driving performance and cognitive workload. There are two main reasons that contribute to this development. First, the amount of time people spend in their vehicles has been steadily increasing, with 86.1% of American citizens commuting in a car, truck or van in 2009 and spending on average 25.1 minutes driving to work (one way) daily, compared to just under 22 minutes in 1980 [1]. And second, with the proliferation of computers and the expansion of communication networks, new types of electronic devices are becoming available and being introduced in vehicles at a rate never seen before [2]. Since driving is usually a monotonous activity (especially everyday commutes to work on familiar roads), those new devices help drivers make their driving experience more interesting and enjoyable. For instance, using a cell phone, smart phone or PDA drivers can send text messages, obtain travel directions, check email, surf the Internet, play hand-held games, and so on. Furthermore, there is plethora of non-hand-held devices, typical examples being car stereos, dashboard GPS units, infotainment systems, and air-conditioning controls, to name just a few. This trend, while certainly exciting and benefiting many areas of our

daily lives, comes at a price of an increased number of accidents caused by driver distraction and inattention [3-7]. For example, based on the results from a naturalistic study, Klauer et al. [4] report that dialing on a hand-held device while driving increases the risk of an accident by a factor of 3.

Very often car manufacturers introduce new safety systems, which are intended to improve driving safety, such as ABS, automatic cruise control, lane departure warnings, etc. Additionally, user interfaces for in-vehicle devices are also changing in order to make interactions relatively safe: hands free phones, speech commands for controlling various devices, and so on. Even though risk homeostasis may be present [8], statistics show that the overall number of car accidents keeps decreasing. Based on a NHTSA study [5] published in 2010, the overall number of crashes decreased from 39,252 in 2005 to 30,797 in 2009. However, according to the same study, the percent of crashes which were associated with driver distraction increased from 10% to 16% for the same 5-year period. Furthermore, the percent of fatalities with reported driver distraction also increased from 10% to 16%. These are important facts which demonstrate how pressing the issue of driver distraction is. Hence, it is of the utmost importance to have reliable tools to detect the potential for distraction that an in-vehicle device has before it is introduced in vehicles.

The facts outlined in the previous paragraphs are not too surprising, since driving itself is a fairly involving activity which requires a complex interaction between both mind and body. Given that every task involves reasoning (possible exceptions being those relying upon muscle memory), the emphasis here should mostly be on the mental activity. Each task has a set of expectations associated to it with respect to the quality of

the performance [9]. Often times it is the case that the expectations are not met despite the individual's ability and motivation to perform the task according to expectations. These failures in performance indicate increased difficulty of the task and the individual's inability to cope with that increase. This gives rise to the concept of increased cognitive load (or workload, which will be used interchangeably in this dissertation). A common definition of cognitive load is the amount of demand which is imposed on an operator's limited mental resources as a result of engagement in a task [10;11]. If we apply this definition to the driving domain it implies that by introducing side tasks drivers have to share their cognitive capacity between driving and side tasks. This may draw attention away from driving, which can lead to accidents.

There exist various measures which reflect changes in cognitive load and can be divided into three general groups: performance-based (usually driving performance measures in the automotive context), physiological and subjective. Each of these groups has a wide variety of measures that are used for estimating the influences of various in-vehicle devices on cognitive load, but some of the more popular ones are as follows:

1. driving performance measures [12-25]: lane position, longitudinal and lateral velocity, steering wheel angle, following distance, acceleration, etc.,
2. physiological measures [12;15;23;24;26-34]: percent time drivers spent looking at the road ahead, changes in gaze location, heart rate, heart rate variability, skin conductance, pupil diameter, respiration, etc.,
3. subjective measures [12-14;23;24;27;28;35]: post experiment questionnaires and rating scales for assessing usability and the level of distraction.

A plethora of studies show that none of the above measures is a panacea. As Wickens [10] points out, we need multiple measures converging in the same direction in order to avoid circular arguments, such as “a task interferes more because of its higher resource demand, and its resource demand is inferred to be higher because of its greater interference.” Furthermore, depending on the experimental conditions, different measures may show different sensitivity. Since many of the above measures were used in the studies presented in this dissertation, a more detailed explanation of their relationship with cognitive load will be provided in Section 2.2.

A common approach to analyzing the influence of an in-vehicle device or interface on driving is to conduct experiments in which participants perform a test drive once with and once without the interface in question (of course, this approach is readily extended to a larger number of experimental conditions). During the experiment various performance measures are collected, such as lane position, steering wheel angle, distance (gap) behind a lead vehicle, and so on. Those measures are then post processed to obtain certain “average-based” measures, such as variances or standard deviations (SD). In general, an increased variance (or SD) of these collected performance measures indicates worse driving performance. Strictly speaking, average values (means) of the above variables can be calculated as well, however, they are often not informative enough. Namely, one can drive close to the edges of the lane throughout the experiment without any negative consequences. What is more informative to look at is how much the position in the lane varies, since it may indicate driver’s higher expanded effort to perform well.

Post-processing calculations of the performance measures usually follow one of two approaches. In one approach, researchers collect values of a desired performance

measure over long stretches of road (i.e. an entire experimental run). Variance (or SD) is then calculated based on all collected data points. A good example may be driving on a straight portion of the road, while continuously interacting with an in-vehicle device [25]. In another approach, the experiment is first divided into multiple segments and the variance (or SD) of a desired performance measure is calculated for each segment individually. Finally, an average of those variances is calculated over all available road segments, possibly weighing each segment's contribution to the average based on the segment length or the time it took to cover the segment. Driving in a city environment with many turns is a good example for segmentation, since the intersections represent natural boundaries between individual streets [23;36]. Whichever approach is selected by the researcher, the same approach is used for each participant and each experimental condition (in-vehicle interface or device on test). Finally, the extracted measures are grouped for each experimental condition separately and analyzed using statistical methods (such as ANOVA and t-test) in order to establish if there are statistically significant differences between the groups. If the differences prove to be significant, it is an indication that the two conditions are not the same and the difference is caused by the experimental condition, given there are no other differences between the two test drives.

1.1 Problem

1.1.1 Example Studies Reporting High Sensitivity of Average-based Measures

The above procedure has proven itself very effective for detecting changes in driving performance caused by ongoing manual-visual interactions. For example,

Salvucci and colleagues [18] examined the impact of MP3 player interactions on driving performance. Specifically, they collected two dependent variables: lateral position deviation (computed as the root-mean-squared error between the center of the vehicle and the center of the lane) and average vehicle speed change. The experimental conditions consisted of normal driving without any interactions with the device (baseline) and three interaction types: selecting and playing songs, podcasts and videos. The experiment was conducted in a simulated highway environment with one lead and one trailing vehicle. Except for playing tasks in case of lateral deviation, both selection and playing tasks significantly impacted each of the two driving performance measures. The authors' overall conclusion was that the tasks that are visually intensive are likely to have detrimental effects on driving, since visual modality is the resource that has to be shared between driving and the side task. Furthermore, we argue that the frequency of the interactions can also play an important role: more frequent interactions (such as with an MP3 player in this study) are likely to influence driving more. Conversely, if the interactions occur infrequently it is possible that their effects on driving may be missed as a result of averaging driving performance measures over time. This suggests that an interaction may still be unsafe, even if our analysis misses it.

One of the most studied effects on driving performance is the one resulting from mobile-phone interactions. Those interactions usually consume considerable amount of time and, at least in the case of hand-held phones, require physical manipulation of the device itself. In both on-road and a driving simulator study Reed and Green [37] investigated the influence of periodically dialing phone numbers using a hand held mobile-phone. Their results demonstrated highly significant effects of the phone task

compared to unencumbered driving on lane-keeping performance (expressed through standard deviation of lane position and steering wheel angle, steering reversal frequency and average lateral speed) under both simulated and on-road conditions.

In one of our early studies [25] (“Interacting with Mobile Radios”), we compared the influence of two interface modalities on driving performance while interacting with police radios. We chose to examine radio interaction for two reasons. First, the radio is one of the most frequently used devices in the police cruiser. Second, interacting with the radio requires taking one’s hand off the wheel and eyes off the road, both of which make crashes more likely. Since police radios have hundreds of channels, they are organized into logical groups called zones. Reaching a particular channel requires first selecting the correct zone and then the desired channel. State-of-the-art police radios require officers to use their hands to change zones and channels, which they do by operating hardware buttons on the faceplate of the radio. They also need to look at a display on the faceplate to verify that the correct zone and channel were selected.

Our hypothesis was that interacting with the police radio using a speech user interface (SUI) provided by the Project54 system would introduce a much smaller degradation of driving performance than using an interface that requires manual interaction. Project54 [38] is a software based package that integrates off-the-shelf electronic devices commonly used in police cruisers and enables an officer to control these devices using voice commands. In our driving simulator-based experiment the primary task was driving while following a lead vehicle at a constant speed of 55 MPH and maintaining a constant distance (gap) behind it. The experiment was performed on a straight, three-lane highway road with light traffic in daylight. The secondary task

consisted of changing channels and zones on a police radio and was performed both using the hardware controls installed on the radio faceplate (manual interaction) and using the Project54 SUI (spoken interaction). The experimental setup is depicted in Figure 1.1. In the case of manual interaction, participants used the buttons (zone up/down and channel up/down) and the display on the radio control head. In the case of spoken interaction, participants issued commands to the SUI specifying the desired zone and channel within that zone. For this purpose the participants used a push-to-talk button (PTT) mounted on the steering wheel that had to be pressed while issuing a command. The experimenter prompted participants to change zones and channels verbally providing the zone and channel names.

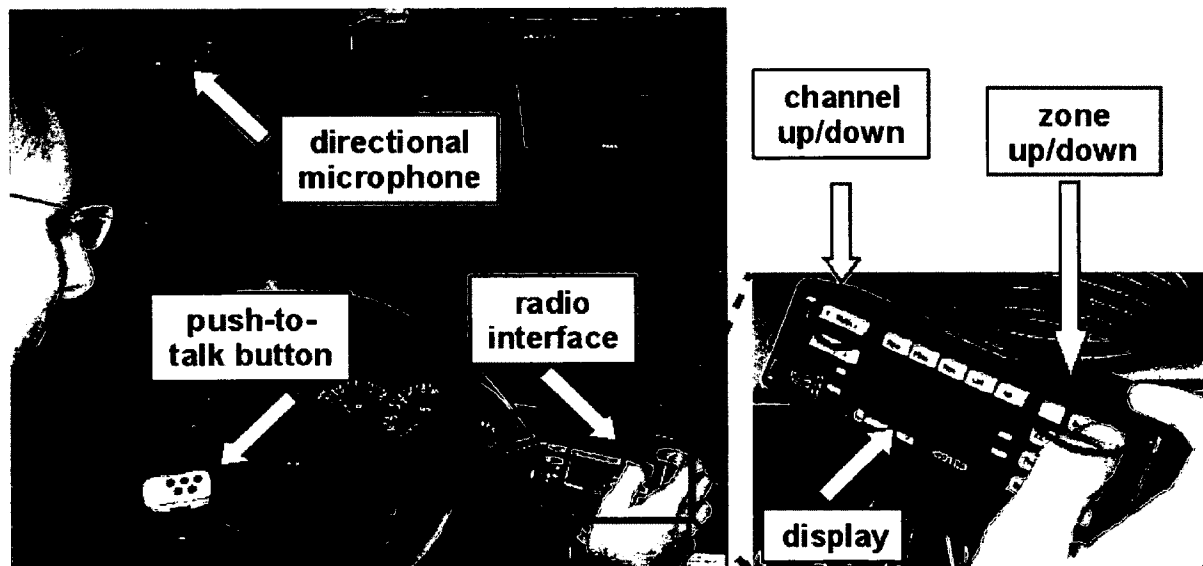


Figure 1.1 Participant manually adjusting channels on the radio inside the simulator.

We estimated driving performance by calculating variances of three dependent variables: velocity, lane position and steering wheel angle. Variances were calculated for the two interaction conditions (manual and spoken interaction) as well as for the baseline condition when the participants were just driving without any distractions.

We found no statistically significant difference between variances for data collected under the baseline conditions and during spoken interactions. However, there was a highly significant effect of the task condition (manual vs. SUI) on the variability of all dependent variables: velocity ($p=0.00035$), car lane position ($p<0.0001$), and steering wheel angle ($p<0.0001$). Box-plots of variances of all dependent variables for all participants and both task conditions (manual and SUI) are shown in Figure 1.2.

One explanation for the above results is that the manual interaction with the police radio required releasing the steering wheel and at the same time looking away from the road (which can be clearly seen in Figure 1.1), and this had a detrimental effect on driving performance. In this experiment we did not collect eye-tracker data, which prevents us from precisely quantifying the amount of visual distraction involved with interactions. Nevertheless, we can qualitatively say that the visual attention to the road was higher in case of SUI interaction.

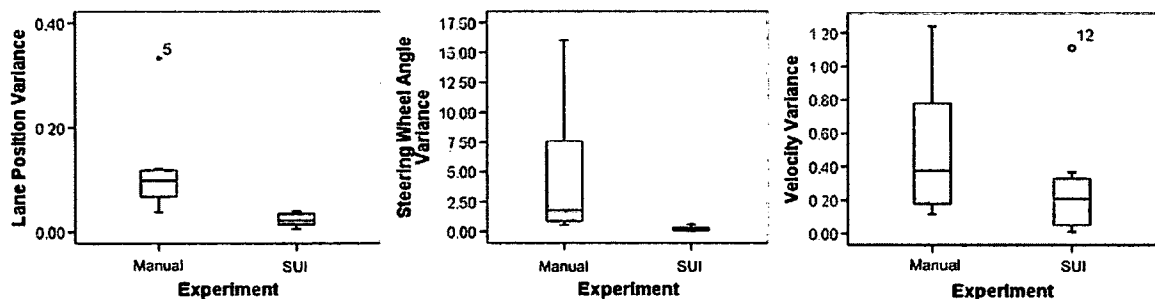


Figure 1.2 Box plots of variances for lane position, steering wheel angle and velocity.

For most participants changing channels manually resulted in drastic changes in driving performance between the baseline and the manual interaction task condition that could be observed even by just plotting the time graphs for the dependent variables. As an example, raw lane position data, recorded for one of the participants, during the

manual (left graph) and speech (right graph) interaction experiment is depicted in Figure 1.3. In both graphs the period until about 130 seconds represents the baseline driving without any interactions. In the case of manual interaction, the vertical dotted lines represent the instants in time when the participant pressed a button on the radio control head. In the case of speech interaction, the dotted lines represent the beginnings of spoken interactions (issuing voice commands). By visually comparing these graphs we can say that the speech interaction introduced little if any additional variation of the lane position, while the manual interaction did.

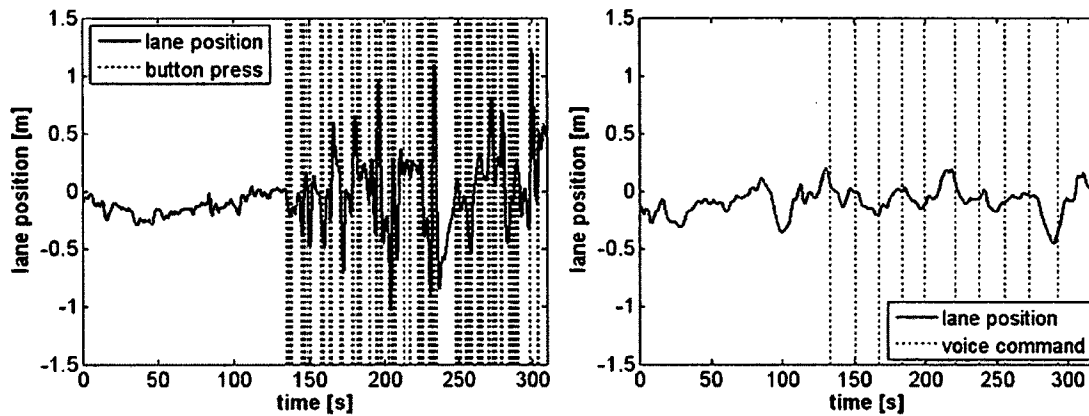


Figure 1.3 Comparison of lane position for manual (left) and speech interaction (right) for one example participant.

The results obtained through the driving performance measures were also reflected in the subjective estimates of workload using the NASA-TLX questionnaire. All participants reported that they experienced a significantly higher workload ($p=0.002$) during manual interaction as is depicted in Figure 1.4.

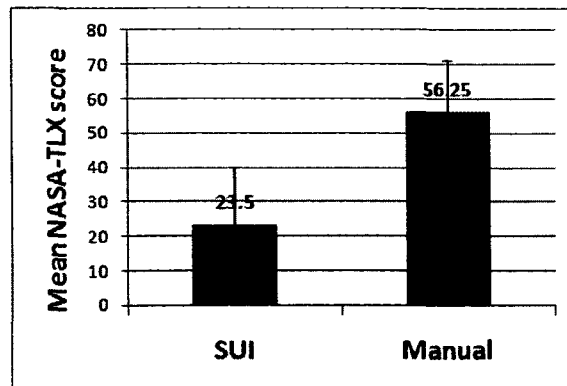


Figure 1.4 Mean NASA-TLX workload score (error bars represent ± 1 SD).

In order to gain better understanding of the effects of speech user interface characteristics on driving performance, we conducted a follow up study [22] (“Speech Interface Accuracy and Driving Performance”). Namely, we examined the effects of three SUI characteristics on driving performance: speech recognition accuracy, PTT button usage and dialogue repair. Speech recognition accuracy was a within-subjects variable and it had two levels: high (89%) and low (49%). Both accuracies were fixed using the Wizard-of-Oz approach (we used prerecorded responses, rather than the actual speech recognizer). PTT button usage was also a within-subjects variable with two levels: PTT mounted on the center console and ambient recognition without the PTT button. Finally, dialogue repair was a between-subjects variable and it represented the system’s responses in case of wrong recognitions: for one group of participants the system uttered an incorrect command (misunderstanding), while for the other the system uttered “unrecognized” (no understanding).

The main task was to follow a single lead vehicle on a two-lane, curvy, rural road in daylight with no ambient traffic. Since the simulated road contained many curvy sections, it forced participants to actively pay attention to the driving task, instead of just focusing on the spoken task. The secondary task (spoken task) included changing

channels and initiating message transmissions on a police radio. The participants were instructed verbally by the system where (which Zone and which Channel) to retransmit each of the messages. Thus, a participant would first choose a desired zone (using the “Zone <name>” command), then choose a desired channel (using the “Channel <name>” command) and finally initiate retransmission (using the “Retransmit” command). Following participant’s commands, the system would verbally confirm the selection. After the system confirmed a successful retransmission, the participant had to return to the initial zone (“Zone A Adam”) and the initial channel (“Channel Troop A”). If a command was unsuccessfully recognized (which was judged by the system’s verbal confirmation), in case of misrecognitions, the participants would respond with “Cancel” and issue the correct command again; in case of non-recognitions, the participants would simply reissue the correct command.

Three dependent variables were collected: lane position, steering wheel angle and velocity. Each variable was transformed using variances. A repeated-measures, multivariate analysis of variance (MANOVA) indicated a significant main effect of recognition accuracy on overall driving performance ($p=0.001$), but not of PTT or dialog repair. Furthermore, a significant interaction between recognition accuracy and PTT was observed ($p=0.01$).

To follow up on the significant effects, we analyzed the effects on each driving performance measure individually using a univariate ANOVA. We found that the recognition accuracy significantly impacted steering wheel angle ($p<0.001$), but not lane position or velocity. This indicates that when the speech recognition accuracy was low, the participants invested more effort to keep the vehicle in the lane. Furthermore, we

discovered that there was a significant effect on lane position of the interaction between recognition accuracy and PTT usage ($p < 0.05$). Namely, when the recognition accuracy was low and when the PTT button was used, lane position variance increased. Conversely, when the recognition accuracy was high, the usage of the PTT button did not affect lane position. No effects were found for steering wheel angle and velocity. Figure 1.5 shows the mean steering wheel angle variance for two recognition accuracies (left graph) and mean lane position variance depending on the PTT usage (right graph).

Eye-tracker data was not collected in this study. However, since the voice commands were used in each case, we can qualitatively say that the visual attention to the road ahead was high. Furthermore, based on the interaction types, we can qualitatively say that the cognitive load was higher in case of low recognition accuracy compared to high recognition accuracy. Even though the number of issued voice commands between the low and high recognition accuracy conditions was approximately the same, the fact that the low accuracy required reissuing voice commands made interactions more difficult and possibly increased participants' frustration. This had a substantial influence on driving performance and was successfully detected by the average-based measures.

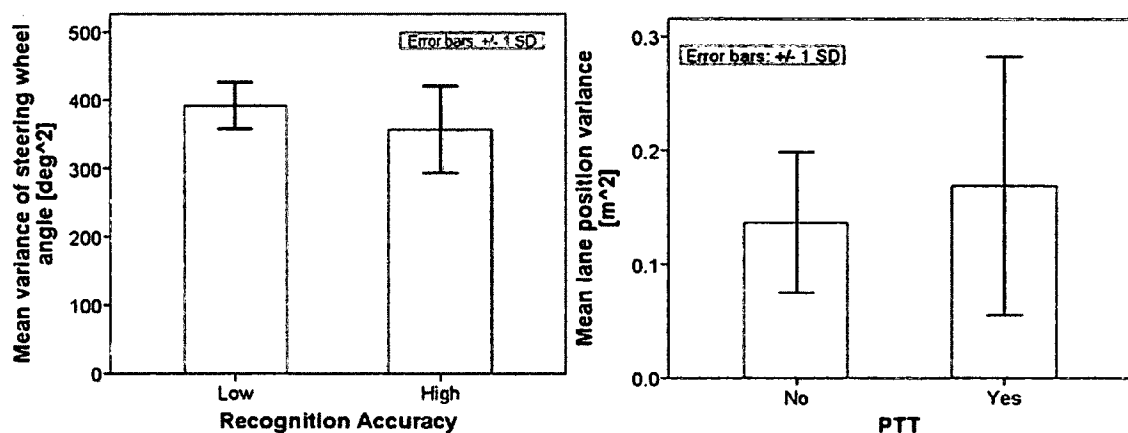


Figure 1.5 Mean steering wheel angle (left) and lane position (right) variances.

1.1.2 Example Studies Reporting Low Sensitivity of Average-based Measures

In the studies presented so far the effects of interactions were long lasting, so their influence on driving could be detected using average-based (mean, standard deviation, and variance) performance measures over time. However, the *problem* is that this approach may not be adequate in all cases. For example, a study by Ranney et al. [39] investigated the influence of various secondary tasks on driving accomplished using either a manual-visual or voice interface. Besides driving-only (no interactions), there were three secondary tasks in total: baseline (continuous phone dialing or radio tuning), simple (searching for a specified message and recoding a voice memo) and complex (same as the simple tasks with the addition of finding and dialing a phone number and retrieving information from an automated phone system). All tasks were performed on a test track while following a lead vehicle.

Among others, the results indicate no difference between the two interfaces (manual-visual vs. voice) regarding following distance to the lead vehicle, no difference regarding the number of steering reversals per second and a significant difference regarding standard deviation of lane position. If we look at the influence of secondary task type, for all of the above variables, driving-only produced significantly smaller effects, while no differences were observed between other secondary tasks (baseline, simple and complex). This result was unexpected, since the complexity of the tasks was very different, which is corroborated by observed significant differences in task completion time. The lack of difference between the complex and the other tasks was especially intriguing, given the increased difficulty of the complex task reflected in the

number of procedural steps and memory burden. The authors suggest that this result may be due to drivers finding a temporary relief from the increased workload during the phone call connect time. This indicates that driving performance deteriorated while performing a task and then improved during the call connect time. However, since the average-based measures characterize each segment as a whole, it is impossible to isolate just the effects of interactions. All dependent variables were collected during two straight portions of the test track, each being 2 miles long. If we take into account that the lead vehicle's average velocity was 40 MPH, it can be calculated that the duration of each segment was about 180 seconds. Average task completion times for easy, baseline and complex interactions were 69.9, 117 and 148.3 seconds, respectively. If we consider these long completion times it is even more curious that no differences have been observed between interaction types. It may be the case that the average-based driving performance measures "smeared" the effects of individual interactions thus preventing us from seeing the changes in driving between these markedly different interaction types.

Another good example where average-based measures may not work well is the interaction with a personal navigation device (PND). This kind of interaction is often not an ongoing activity: drivers might look at a PND map for several seconds, but do this infrequently. Additionally, not every glance at an in-vehicle display results in worse driving performance. In these cases averaged performance measures might not adequately capture the negative influence of the interaction on driving, since driving performance deterioration occurs for relatively short periods of time compared to the duration of the experiment or segment.

In a navigation study [40] (“The Effects of PNDs on Driving and Visual Attention”) we analyzed how driving performance and visual attention change as a result of three navigation alternatives: printed paper directions, standard map-based directions and voice-only directions. Printed paper directions served as a baseline. Even though PNDs are very common in vehicles nowadays, some drivers still use paper directions, which is why we decided to include them in this study. The participants were provided with printed directions similar to those that can be obtained from popular web services: map of the route and a list of turn-by-turn directions. This navigation required a manual-visual interaction, since the participants had to handle the sheet of paper with their hands. Standard map-based directions (SPND) simulated commercially available PNDs and provided a map with a real-time location of the vehicle (green triangle in Figure 1.6) as well as verbal prompts for the upcoming turns. The map also contained an outline of the route to be traversed (solid red line in Figure 1.6). Spoken directions (voice-only) were included in the study in order to investigate whether the visual presentation of directions on standard PNDs negatively influences driving and visual attention. Therefore, we used the same spoken directions as with the SPND, except that the map was not visible.



Figure 1.6 7" LCD screen simulating a map-based standard PND.

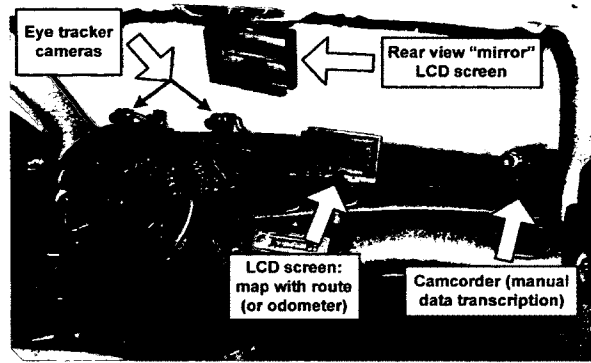


Figure 1.7 Experimental setup inside the simulator cabin.

Figure 1.7 shows the equipment setup inside the simulator cabin. For this study we used an eye-tracker, which enabled us to precisely quantify the amount of visual attention the participants directed to the road ahead. In this figure we can also see the location of a 7" LCD screen which simulated standard PND directions. Unless a PND is already embedded in the center column, drivers typically mount PNDs on the windshield or on top of the dashboard. In this study we decided to place the LCD screen on top of the dashboard, since this location requires smaller gaze changes compared to a screen which is integrated into the dashboard.

In our driving simulator-based experiment the main task was to navigate through a simulated environment using the above navigation devices. The simulated environment consisted of multiple road types; however, we decided to process the data from two-lane city roads with lane markings. This ensured that the characteristics of all selected road segments were the same. The path included multiple left and right turns, so we segmented the experiment such that each individual street was considered as a separate segment. The intersections were used as the natural boundaries between the segments. Furthermore, we excluded the data from the intersections, because the variances resulting from the turning maneuvers are much higher than the variances

encountered while driving on straight segments, which would likely mask any device effects. We collected multiple dependent variables for each segment: variances of lane position, steering wheel angle and velocity, average velocity and percent dwell time (PDT) on the outside world. Since the segments were of different lengths, we weighted the contribution of each segment to each driving performance dependent variable based on the ratio of the time each participant spent on that segment to the overall time spent on all segments together.

By conducting one-way ANOVAs for each dependent variable we obtained significant main effects of navigation type on the following dependent variables: variance of lane position ($F(2,20)=4.94$, $p<0.05$), steering wheel angle variance ($F(2,20)=4.67$, $p<0.05$) and PDT on the outside world ($F(2,20)=14.03$, $p<0.001$). No significant effects were observed regarding the velocity variance or average velocity.

Figure 1.8 shows the mean lane position variances (left) and mean steering wheel angle variances (right) for the three navigation aids. For both dependent variables, post-hoc pairwise comparisons indicated significant differences between paper directions ($p<0.05$) and both SPND and voice-only directions. No differences were observed between SPND and voice-only directions. The results obtained using the lane position variance indicate that when paper directions were used, participants were unable to control the position of the car with the same degree of accuracy as with the other navigation aids. Similarly, steering wheel angle variance indicates that participants invested a significantly higher effort on steering when paper directions were used.

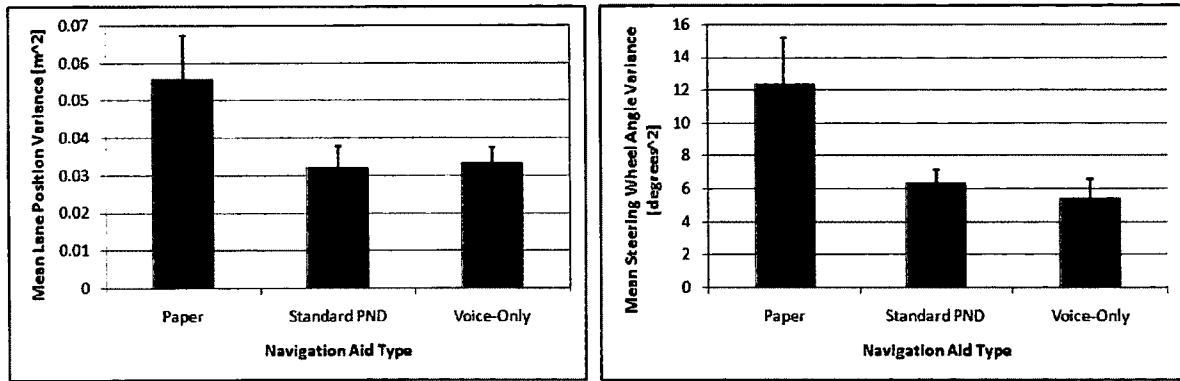


Figure 1.8 Mean variances of lane position (left) and steering wheel angle (right).

Figure 1.9 shows the average PDT on the outside world for the three navigation aids. Not surprisingly, participants spent the least amount of time looking at the forward road when they used the paper directions. This was corroborated through the post-hoc comparisons, which indicated that paper directions caused the smallest PDT, followed by SPND and voice-only navigation aids. All pairwise comparisons indicated statistically significant differences ($p < 0.05$).

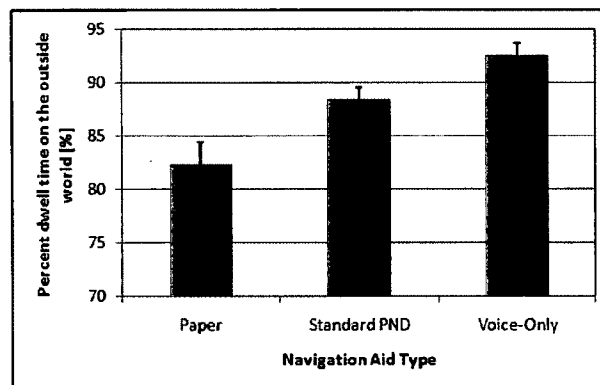


Figure 1.9 PDT on the outside world.

One explanation for these results is that the paper directions required manipulating a sheet of paper and on average caused longer glances (1.4 sec) compared to SPND (0.6 sec). This affected driving performance substantially, thus enabling us to see the influence on driving performance measures using the averaging approach.

Conversely, averaging did not uncover differences between SPND and voice-only PNDs, despite the significant difference ($p < 0.05$) in visual attention between the two (88% vs. 92%, respectively). However, the fact that we did not find any significant differences in driving performance does not mean that there is none – merely that, with our simple driving task, the null hypothesis that the driving performance when using standard PND directions is the same as that when using spoken directions only could not be rejected.

Similar results were obtained in a follow-up study [23] (“Glancing at PNDs Can Affect Driving“) which was intended to investigate more closely the impacts on driving performance produced by standard PND and spoken directions only. The simulated scenario was more challenging than the previous one and it resembled a two-lane city road, which was populated with realistic traffic, pedestrians and unexpected events (cars braking, pedestrians jaywalking, etc.). This substantially increased the level of realism, since now the participants actually had to pay close attention to the virtual world, specifically cars and people. Figure 1.10 shows how the simulated road looked like. It also shows one of the unexpected events that occurred during the experiment: a pedestrian emerging from behind a parked vehicle in front of the participant’s vehicle.



Figure 1.10 Simulated two-lane city road with unexpected event.

We collected multiple dependent variables: variances of lane position, steering wheel angle and velocity, average velocity, number of collisions with other objects (pedestrians and cars) and PDT on the outside world.

The navigation route included many streets that the participants were supposed to traverse. Similar to the previous study, we segmented the experiment using the intersections as the natural boundaries. In this experiment we focused on 13 segments for which we extracted all of our dependent variables. Each segment was 200 meters long and all had the same characteristics. Segments where unexpected events occurred were excluded from the analysis, since an unexpected event may require sudden breaking and/or steering wheel motion, which can impact driving performance significantly, thus making comparisons with other segments difficult.

After performing a one-way ANOVA using PDT as the dependent variable, we found a significant main effect of the navigation type on visual attention ($p < 0.01$). As expected, time spent looking at the outside world was significantly higher in case of spoken directions (96.9%) compared to SPND (90.4%). These results are in agreement with the ones obtained in the previous study.

Figure 1.11 shows changes in PDT on the outside world (left) and PDT on SPND (right) based on the distance from the previous intersection. We can see that the participants were more likely to look at the PND right after making a turn. This can be explained by the drivers' urge to confirm whether they made a correct turn as well as the need to observe the upcoming direction. Furthermore, participants were less likely to look at the PND as they approached the next intersection, which indicates that they were focusing more on becoming ready to make the upcoming turn.

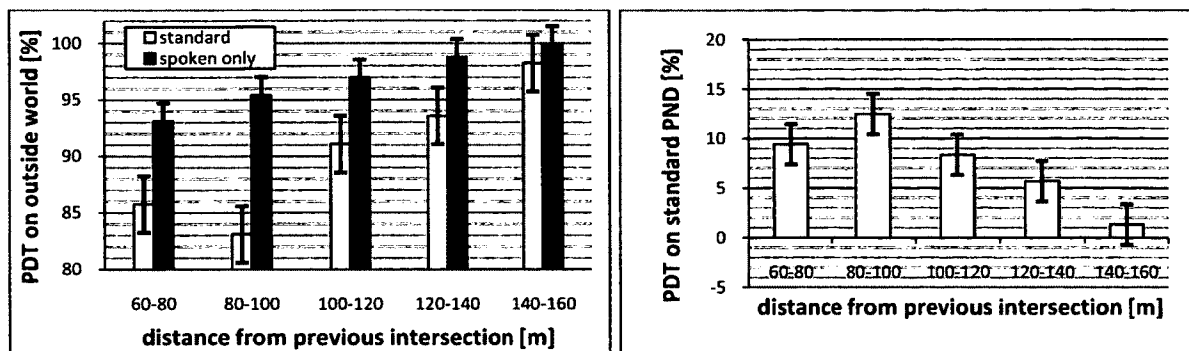


Figure 1.11 Changes in PDT on the outside world (left) and SPND (right) based on distances from intersections.

We found no significant differences between the navigation devices in any of the average-based driving performance measures, even though arguably large significant differences in visual attention were detected. Specifically, participants on average spent about 6.5% more time looking at the road ahead when using the spoken output-only PND – a difference of about 4 seconds for every minute of driving. The lack of the observed effects on driving agrees with the findings from the previous study and is equally surprising given the impacted visual attention.

The results from these two navigation studies suggest that the glances directed towards the PND displays have short-lived, local influences which are easily lost in the averages. As we can see from the graphs displayed in Figure 1.11, visual attention to the road ahead varies widely depending on the car's physical location within each street segment. It is likely that the negative influences of looking away from the road were localized in the areas of the segment where the drivers directed their visual attention away from the road the most. In this case, the largest difference in PDT directed to the road between spoken-only and standard PND occurred between 80 and 100 meters from the beginning of the segment and is equal to 12.31%. However, since the visual attention

through the rest of the segment was not impacted severely, it can be expected that the driving performance was satisfactory. Thus, in the process of averaging over the duration of the experiment, the predominantly satisfactory driving performance overwhelms possible short-term deteriorations. As a consequence, statistically significant differences between conditions that involve such interactions may not be established. Figure 1.11 also indicates that the observed overall difference in visual attention (PDT) does not provide the complete picture about the way participants interact with in-vehicle devices.

Nevertheless, just because the influence of in-car interactions does not show in statistical analyses of long periods, it does not mean that it should be neglected, especially if the nature of the interaction is such that it can be performed very often. This assertion is corroborated by a naturalistic study done by Klauer and colleagues [4] who obtained a correlation coefficient of 0.72 between the frequency of drivers' involvement in inattention-related tasks and the frequency of being involved in inattention-related crashes and near-crashes. Furthermore, they calculated that the odds ratio of being involved in a crash or near-crash even for simple interactions (such as adjusting a radio, talking to a passenger, drinking, etc.) is 1.18, while for complex interactions (such as dialing a hand-held phone, operating a PDA, etc.) is 3.1. These are all very important implications that should be accounted for when analyzing in-vehicle interactions. We can also argue that human psychology is a factor that plays a very important role in driving environment. Driving is a forgettable activity, which means that the importance of previous incidents decays over time in drivers' minds. Thus, it is possible that a driver may engage in the same activity again after a long enough time. Additionally, if the risk

of interactions is within the subjective threshold, the engagement may be continuous (good examples being a cell phone conversation or an MP3 player interaction).

The averaging problem observed in the previous studies may sometimes occur with manual-visual interactions as well. Hosking et al. [34] conducted a driving simulator study which was intended to investigate the impacts of sending and retrieving text messages using a cell phone on driving performance and visual attention of young novice drivers. Retrieving was defined as opening and reading a text message, while sending was defined as writing a reply to a text message and sending it. The simulated environment consisted of a two lane city road with multiple critical events: stopping at a red light initiated at the predefined distance from the signal, three car following tasks where the driver had to maintain safe distance (gap) behind a lead vehicle, two lane changing tasks where the driver was changing lanes according to signs located at the side of the road, avoiding a pedestrian and avoiding an oncoming vehicle which was turning in front of the participant. Each of these tasks was completed under both text messaging (retrieving + sending) and non-text messaging conditions, where the latter was used as a control. Multiple dependent variables were collected: averages and standard deviations of lateral position and speed as well as the proportion of time spent not looking at the road (equivalent to PDT off road).

Two sets of results were obtained depending on the way data was analyzed. In the first case, the data was aggregated across all events for the time periods corresponding to retrieving and sending text messages. The results indicated a significantly larger proportion of time not looking at the road ($\approx 40\%$) in case of text messaging (retrieving + sending) compared to non-text messaging condition ($\approx 10\%$).

However, no differences were revealed regarding averages and standard deviations of either lateral position or speed between the two conditions for both sending and retrieving time intervals. This was indeed unexpected given the large observed difference ($\approx 30\%$) in the visual attention directed off-road. The results are somewhat different when each event is analyzed individually. Namely, standard deviation of lateral position was significantly higher during sending time intervals (compared to non-text conditions) for three out of eight events: avoiding a pedestrian, red light signal and the second car following event. However, no differences were observed during retrieving time intervals for any of the driving variables. The lack of difference between retrieving time intervals and non-text conditions is even more unexpected if we look at the subjective assessments, which indicate that 95% of participants reported that their driving performance declined when receiving messages. It is likely that in this study average-based measures were not sensitive enough to isolate the effects of cell-phone interactions from the effects caused by critical events.

1.1.3 Problem Overview

Figure 1.12 presents one specific example obtained from a driving simulator which illustrates the averaging problem visually. The upper graph shows the lane position signal divided into two regions: “interaction” region where the participant was interacting with an iPod and “just driving” region where the participant did not perform any side tasks. Both regions are about 20 seconds long. The lower graph shows where the driver’s visual attention was directed to over time: 1 indicates speedometer, 5 indicates looking at the road ahead and 8 indicates looking at the iPod. We can clearly see that the participant drifted towards the edge of the lane while looking at the iPod and then brought the car

back to the original location (about -0.2 meters) after returning the gaze back to the road. After performing the calculations, we can see that both regions have very similar average values ($\bar{x}_{Interaction} = -0.09$, $\bar{x}_{Just\ driving} = -0.12$) and standard deviations ($s_{Interaction} = 0.032$, $s_{Just\ driving} = 0.012$). Even though this is a very simplified example, after performing a two sample t-test between the two regions of lane position, we obtained a p-value of 0.1631, which indicates that there is no difference between the two signals. However, we can clearly see that something actually did happen during the “interaction” region at about 388 seconds. Namely, the participant drifted for more than 0.5 meters as a result of interacting with an iPod. This simple example demonstrates how the short lived, but nevertheless important events, may get “washed-away” in the averages.

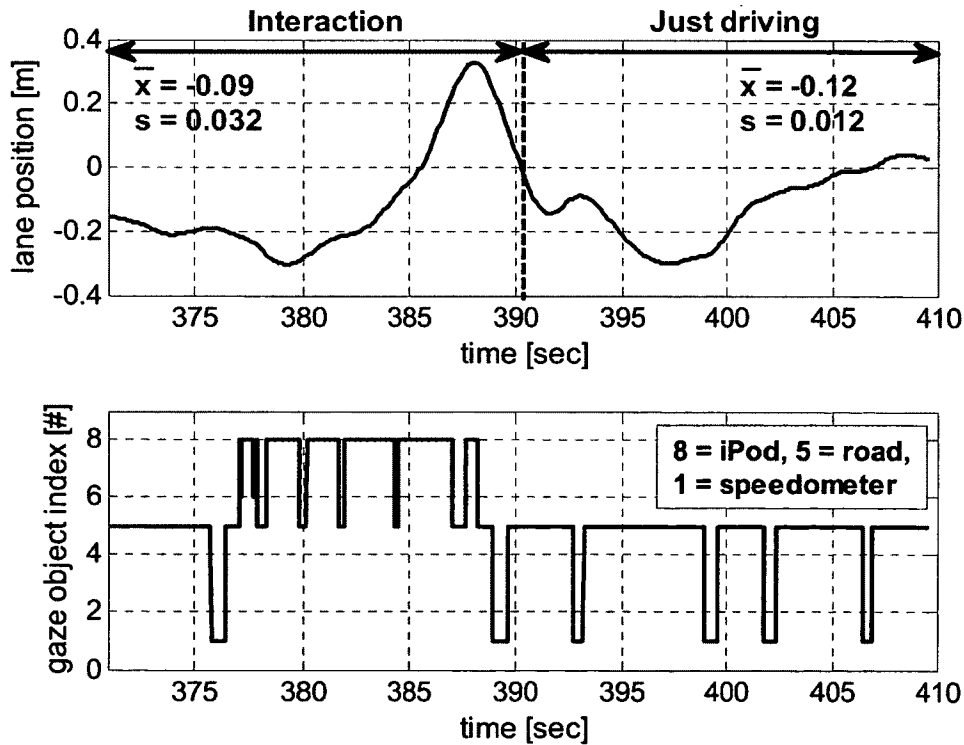


Figure 1.12 Problem illustration using a specific example.

Before concluding this section we will make a simple thought experiment in order to shed light on the above problem from a real-life perspective. Imagine a situation where a driver is using a PND to navigate on an unfamiliar road. The PND informs a driver to make a right turn in 0.1 miles. However, there are multiple right turns in close proximity and the driver decides to glance at the PND in order to decide which one to take. While looking at the PND, the driver drifts towards the sidewalk. After returning the gaze to the road the driver notices that the car is very close to a pedestrian standing at the curb, so she executes a correction maneuver to re-center the car in the lane. The rest of the trip goes without any incidents and the driver arrives successfully at the destination. If we look at the above drive from a high-level perspective, it was a successful one, since no collisions occurred. Similarly, if we look at some more specific descriptors of driving performance, such as the average values or variances of lane position and steering wheel angle, it is likely that no differences will be detected, compared to a similar drive without any incidents whatsoever. The reason for this is that the duration of the incident was short-lived, thus producing a small impact on the long-term average. While it is obvious that looking at the PND affected driving in this example, average-based driving performance measures do not capture the obviousness of this situation. Furthermore, incidents of this type happen fairly infrequently.

The studies presented in this section sample the space of in-vehicle interactions fairly well, from manual-visual and spoken interactions to purely visual interactions. Based on these results we can conclude that there are three main aspects of the problem:

1. As was demonstrated through the previous studies we are often unable to observe changes in cognitive load through differences in average-based driving

performance measures, such as lane position and steering wheel angle, even though differences may exist in visual attention and/or subjective estimates. Nevertheless, localized changes in performance measures may still exist, which indicate an effect of side task engagement, as was the case in Figure 1.12. There are three main effects that may contribute to missing localized changes: observation intervals are significantly longer than the duration of the localized change, localized changes occur infrequently and non-interaction related changes in driving performance may mask relevant changes coming from in-vehicle interactions. Therefore, a performance measure which would be able to account for such cases is currently not available.

2. The second aspect of the problem is the inability to demonstrate changes in cognitive load from multiple sources. Namely, in our navigation studies we observed highly significant differences in visual attention to the forward road. This indicates changes in cognitive load. However, this way we possess only a single evidence pointing to that conclusion, which can then lead to a circular argument: visual attention to the road is low thus cognitive load is high, and cognitive load is high because visual attention to the road is low. A much stronger argument could be made if we would possess an additional measure suggesting the same conclusion.
3. Finally, a coarse measure, such as the number of collisions, may be useful in characterizing the overall risk of using a particular device. However, collisions occur very rarely. Thus, a finer measure is required in order to facilitate the design process.

1.2 Goals

Motivated by the above problem, we can state that the goals of this dissertation are as follows:

1. *(G1) Introduce a cumulative measure of a secondary task engagement on cognitive load.* This measure should tell us how cognitive load is influenced over the course of performing the secondary task. It is understood that cognitive load is not constant. Rather, during periods of engagement in the secondary task cognitive load is increased. When there is no engagement in the secondary task, cognitive load is reduced. Our goal is to create a single measure that reflects both the impact of the periodic engagements in the secondary task (e.g. a driver glancing off the road from time to time, in order to look at the map of an in-vehicle navigation device) and the frequency at which this activity happens in order to accomplish the secondary task (e.g. to navigate from point A to point B). Such a measure would provide more sensitivity to cognitive load changes compared to standard average-based driving performance measures. We also require this measure to allow ranking of the results obtained for different types of secondary task engagement, which can then be used to compare different designs.
2. *(G2) Introduce an instance-based measure of a secondary task engagement on cognitive load.* This measure should tell us how cognitive load is influenced, on average, by an instance of engagement in the activity (e.g. how does one glance at the map influence cognitive load, on average). This approach complements the cumulative findings by allowing low level insight into individual engagements. Similar to the first goal, we also expect this measure to allow ranking of the

observed results. This is very important, because it allows comparing the effects of different interaction types at the level of *individual* secondary task engagements.

3. (G3) *Provide explanation for the mechanisms underlying the cumulative and instance-based measures.* Knowing the underlying mechanisms has two advantages. First, it allows us to propose explanations about why the results behave in the observed fashion. And second, it gives us the ability to foresee what the results may be in advance, which may be used to inform design decisions. For example, if we are given a choice between multiple interaction modalities with a particular device, the obtained mechanisms may help us in ranking these modalities with respect to their impact on driving and cognitive load.

1.3 Hypotheses

Based on the results obtained from the previous studies we can say that the average-based driving performance measures do not characterize the potential causes of the observed changes. In other words, they characterize the experiment (or the corresponding experimental segments) as a whole without regard to when an influence has occurred or what caused it. This exactly leads to the general problem we are addressing: localized changes may be missed in the averages. Therefore, we need a performance measure which would take into account not just the final manifestation of an in-vehicle interaction (such as the effects on lane position or steering wheel angle), but also the potential causes. This agrees with the requirements of the first two goals specified in the previous section.

Generalization

Ideally, this measure should be sensitive to many different interaction types, such as haptic (based on the sense of touch), spoken (speech production and comprehension), olfactory (based on the sense of smell). As we will see later, our method has the potential to be readily extended to the above interaction types as well, which provides generalization. However, as a first step, we will constrain this research to the interfaces that rely primarily on visual and manual-visual interactions. We have to note here that this restriction is not a limiting factor, since visual and manual-visual interactions are two types most commonly used with in-vehicle interfaces [41].

Construct Validity

Another important aspect that this measure should satisfy is construct validity. Construct validity refers to the ability of a specific tool to measure the construct of interest [42] – in our case changes in cognitive load in general and driving performance in particular. As we will see in the following sections, three driving simulator studies will be proposed for testing our hypotheses. These studies will also be used to test construct validity by comparing the results obtained through our method with the results of measures known to be sensitive to cognitive load changes, specifically, average-based driving performance measures (variances of lane position and steering wheel angle), subjective estimates of cognitive load (NASA-TLX questionnaire) and physiological measures (average heart rate and skin conductance). This way we can test whether our method provides conclusions in the same direction as the “standardized” measures. If this proves to be the case, it will be an indication that construct validity is supported.

Proposed Hypotheses

If a driver is actively paying attention to the road ahead, any observed changes in driving performance can be attributed to willful actions. However, while performing side tasks the driver is distracted from the primary task of driving and any observed changes are likely caused by the interactions with the side tasks. Thus, there are two variables of interest here: interaction variable (ρ) which serves as an “initiator” and driving performance variable (θ) which reflects the outcomes of the interactions. To generalize the approach, both of these variables can be transformed using some appropriate functions, $f(\rho)$ and $g(\theta)$. The purpose of the transformations is to filter the raw data in ρ and θ variables in order to emphasize desired effects. We can use $f(\rho)$ to determine when/where the important influences occur and then use that information to extract the effects observed in $g(\theta)$. The approach of extracting relevant information from $g(\theta)$ using the initiator sequence $f(\rho)$ can be defined as follows: $L(f(\rho), g(\theta))$. L can be termed as the “extraction function” as it extracts changes in driving performance initiated by the interactions characterized by $f(\rho)$. Based on these definitions, we can formulate the following hypotheses relating to the first two goals of this dissertation:

- *(H1) Initiator-based quantification of cumulative secondary task engagement.* In this case L uses an initiator sequence $f(\rho)$ which indicates where individual secondary task engagements occur and use those to calculate the overall effect on driving performance and cognitive load.
- *(H2) Initiator-based quantification of instances of secondary task engagement.* This case is similar to the previous one in the sense that an initiator sequence $f(\rho)$ is also used to detect secondary task engagements. However, L should be

modified such that the result reflects the effects of *individual* instances of engagement.

Both the first and the second goal require our proposed methods to allow ranking the results obtained for different types of secondary task engagement. Since the proposed cumulative and instance-based measures are based on the same underlying “extraction function” L , we can expect that they will provide the results of the similar underlying nature. Therefore, we propose one common ranking procedure (“RP”) addressed by Hypothesis H_{RP} :

- (H_{RP}) *Establishing significant differences between secondary task engagements.*

The results obtained for different types of secondary task engagement using the cumulative and instance-based measures can be compared statistically. If the differences prove to be significant, this information may be used for ranking the size of their effects.

The way we proposed the above hypotheses we have separated the quantification (cumulative and instance-based) and the ranking parts of our first two goals. Therefore, when addressing those goals, we will consider the appropriate pairs of hypotheses in concert: for G1 we will use $H1 + H_{RP}$, while for G2 we will use $H2 + H_{RP}$.

Finally, our last goal (G3) is addressed by Hypothesis H3 as follows:

- ($H3$) *Establishing significant predictors.* By revealing the variables which contribute to the cumulative and instance-based results, we can propose explanations for the underlying mechanisms.

The following sections will give more insight into each of the proposed hypotheses. The discussion of the proposed approaches for testing the hypotheses will follow in Section 1.4 (pg. 42).

1.3.1 Hypothesis H1 – Quantifying Cumulative Secondary Task Engagements

The conclusions of our preliminary study [23] will help in defining the first hypothesis. Namely, according to the results obtained in this study the mathematical function of *cross-correlation* appears to be a good choice for function L , at least regarding short, local influences of glances on PND devices (a detailed discussion of the cross-correlation method will be presented in Chapter 3). The study compared a standard, map-based PND with spoken-only directions (no visual feedback) regarding their impact on driving performance and visual attention, which, as we know, reflect changes in cognitive load. The main task involved driving in a city environment and following navigation directions issued by the two PNDs.

As we had a chance to see in the introduction, the results demonstrated a significant difference between the two PNDs regarding visual attention to the road ahead (PDT was 96.9% for spoken-only and 90.4% for SPND). No difference was observed regarding lane position and steering wheel angle variances, which was surprising given the impacted visual attention. Since the localized influences of in-vehicle interactions may be missed in the averages, we expected that glances directed away from the road may introduce short-term changes in driving performance. In other words, after returning the gaze to the road, drivers may need to apply corrections in order to keep a steady

position in the lane. If this was indeed the case, we expected to see peaks in cross-correlation functions calculated for two driving performance measures (specifically, lane position and steering wheel angle) following the return of the gaze to the road.

A sequence of glances (consisting of objects a driver is looking at) was selected as the logical choice for ρ , while lane position and steering wheel angle were used as θ (a separate θ for each driving measure was created). Raw values of ρ and θ are not very informative, so we transformed those using f and g functions as well. Sequence ρ consists of discrete, nominal values which indicate the objects the glance is directed to over time. These values can be used to produce a sequence which aggregates all glances directed off road: cabin, PND, speedometer, etc. Function f accomplishes that by transforming ρ into instantaneous PDT (IPDT) on the outside world. The IPDT was calculated at a 10 Hz rate by calculating a separate PDT for each consecutive 100 ms window of eye tracker data. Since the eye tracker data was recorded at 60 Hz, we calculated instantaneous PDTs using six eye tracker data samples at a time. Finally, the IPDT was transformed such that a value of 0 represented 100% IPDT (attention fully on the outside world), while a value of 1 represented 0% IPDT (attention directed away from the road).

Function g was intended to capture localized changes in driving performance variables resulting from glances directed off-road. It was implemented by calculating short-term, running variances of lane position and steering wheel angle calculated at a 10 Hz rate for 1 second long windows (i.e., for 10 samples of the given driving performance measure at a time). The choice of 1 second long windows reflects our expectation that the corrections to lane position on straight roads, resulting from relatively large changes in

the steering wheel angle, will take less than 1 second. After calculating the variance for each window, the window is moved by one sample and then the next variance is calculated. Since the sampling frequency is 10 Hz, this amounts to a 90% overlap between the windows. The result of each variance calculation is written at the location of the sample which represents the beginning of each corresponding window.

Finally, $f(\rho)$ and $g(\theta)$ sequences were cross-correlated. Cross-correlation is capable of detecting similarities between two sequences, which can be related to each other either causally or indirectly (through known and unknown mechanisms). In this particular case, similarities are expressed through the glances directed off-road ($f(\rho)$), which are resulting in higher variances in driving performance measures ($g(\theta)$). Figure 1.13 shows two cross-correlation functions, one for each driving performance measure: lane position (R_{lp} , left graph) and steering wheel angle (R_{sw} , right graph). The blue and green lines represent cross-correlation functions obtained for standard and spoken-only PND, respectively. The brown dash-dot lines in both graphs represent the significance level of 0.05, which indicates statistical significance of any peaks larger than this level.

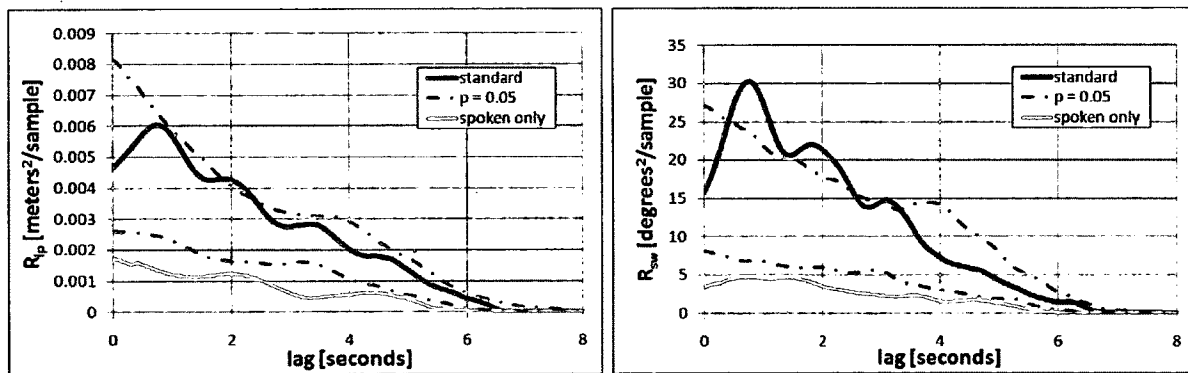


Figure 1.13 Cross-correlation results comparing standard and spoken-only PND as obtained using initial cross-correlation method.

We can see that there is a highly significant prominent peak in case of steering wheel angle cross-correlation function for SPND at the lag of about 0.8 seconds. This lag indicates that an increase in the steering wheel angle variance follows reduced attention to the outside world. The peak also exists for lane position cross-correlation function for SPND at 0.8 seconds; however, it is just below the significance level of 0.05. It is possible that this lack of significance is due to a relatively small number of participants (8) who participated in this study. Nevertheless, this peak indicates an existing trend towards the largest changes in lane position occurring right after returning the gaze to the road. These results suggest that our expectation is clearly supported in case of steering wheel angle, while a trend can be seen in case of lane position. We can also see that there are no significant peaks in case of spoken-only PND. This indicates that participants managed to maintain good control of the vehicle despite occasional glances directed towards the speedometer, dashboard or steering wheel. We also have to notice the large difference between the general levels of the cross-correlation functions obtained for the standard and spoken-only PND. This indicates that SPND introduced higher overall impact on driving compared to spoken-only PND, which can be associated with higher cognitive load.

Based on these results we can say that cross-correlation appears to be a promising choice for function L in hypothesis H1 and may be used to provide a cumulative measure of a secondary task engagement on cognitive load. Nevertheless, it has to be tested under more experimental conditions in order to confirm its usefulness. Furthermore, the previous choice of functions f and g was not ideal, since it resulted in the following difficulties:

- a) The previous method required involved windowing in order to calculate running variances for the driving performance measures. Thus, the result depends on the window duration and the overlap between the windows. Additionally, the direction of the window (whether the window should be applied to the right or to the left from the current location) and therefore the location of the result depend on the windowing. This all causes data smearing. Similar problem exists with the windowing employed for obtaining IPDT from the eye-tracker data. An improvement to this procedure should be devised.
- b) As a consequence of using variances, the units of the results are in $[\text{meters}^2/\text{sample}]$ for lane position and $[\text{degrees}^2/\text{sample}]$ for steering wheel angle, which are difficult to comprehend practically.
- c) Due to the definition of the cross-correlation formula, the result may be skewed towards the long glances, since they contribute more to the overall cross-correlation result. Also, if a glance is observed as a whole, it is unclear which part of the glance is the optimal reference point from which we can measure potential changes in driving performance measures. Therefore, a specific reference point is necessary to make the approach truly initiator-based as indicated in hypothesis H1.

In order to address these issues and by taking into account the results obtained from this study, we hypothesize (H1) that the following choices for L , f and g will be able to satisfy the quantification part of goal G1:

- a) L should be based on the mathematical function of cross-correlation.

- b) f should transform the glance sequence p into a sequence of zeros and ones, where ones represent the instants when the driver returns the gaze towards the road. This way *edges* of the glances represent specific reference points from which we can measure changes in driving performance. This goes along well with our assertion that glances contribute to changes in driving performance.
- c) g should calculate the absolute first difference of the provided driving performance measure (θ), specifically lane position and steering wheel angle. This results in the highest possible resolution without smearing the data, since no windowing is required (strictly speaking, the duration of the window is only two consecutive samples). This transformation is similar to variance in a sense that it resembles the overall change in the data without regard to the direction of the change. The reason for taking the absolute value is that moving too far to either side of the lane of travel produces an equally hazardous situation: vehicles coming from the opposite direction on the left and edge of the road on the right. Similar logic applies to steering wheel angle: pronounced changes to either side indicate potential corrections after returning the gaze back to the road.

1.3.2 Hypothesis H2 – Quantifying Instances of Secondary Task Engagements

The second hypothesis is based on the same main assumption as with H1: glances directed away from the road may produce localized changes in driving performance after returning the gaze back to the road (potential corrective actions). However, in this case we are interested in characterizing the changes in cognitive load

resulting from *individual* interactions (engagements in secondary task). The way the cross-correlation function is defined in H1 provides us with the cumulative effect on cognitive load coming from *all* engagements observed together. Thus, we hypothesize (H2) that the quantification part of G2 can be addressed by introducing a normalization factor to the initiator-based function L . Since in our case (although a generalization is possible to other interaction types and even other domains besides automotive) we consider an individual glance directed off-road as one instance of secondary task engagement, the normalization should be performed with respect to the total number of glances that occurred in the current experimental epoch (segment). Thus, if L is the cross-correlation function, we can define a new extraction function to be $L' = L/N_{glances}$.

1.3.3 Hypothesis H_{RP} – Ranking the Effects of Secondary Task Engagements

Besides quantifying the effects of secondary task engagements, our first two goals are also concerned with ranking the effects of multiple task difficulty levels tested under common experimental conditions. One example would be to make a distinction among several interaction alternatives with a personal navigation device or an MP3 player. We hypothesize (H_{RP}) that the cumulative and instance-based measures based on our cross-correlation method will be able to achieve this goal. Namely, we propose two approaches: magnitudes of the most prominent peaks and areas below the cross-correlation functions. The results obtained for different experimental conditions using these approaches can then be compared statistically. If the significant differences are

observed, we can perform the ranking of the experimental conditions regarding their impact on driving and cognitive load.

1.3.4 Hypothesis H3 – Analyzing Underlying Mechanisms

Our last goal (G3) is to propose an explanation for the mechanisms underlying the cumulative and instance-based measures. We have to note here that our goal is not to obtain a universal model which could be applied to any experimental condition or to demonstrate causal relationship; rather, we intend to reveal a set of variables (in other words predictors) that contribute significantly to the observed results and to demonstrate the predictive ability of our measures.

Both cumulative and instance-based measures are based on the cross-correlation function. In our case the cross-correlation function is applied between the glance sequence and the performance sequence. Therefore, we have to take into consideration various variables that can describe those sequences well. We hypothesize (H3) that the following variables may have an important influence on the observed results: PDT spent looking away from the road, number of glances, average glance duration and average amount of change in lane position, vehicle heading and steering wheel angle.

1.4 Testing Hypotheses

1.4.1 Testing H1 and H2 – Cumulative and Instance-based Quantifications of Secondary Task Engagements

Hypotheses H1 and H2 propose using the cross-correlation function in order to provide a measure capable of estimating cumulative and instance-based effects of secondary-task engagements on cognitive load, respectively.

Since both hypotheses are based on the same underlying assumptions, the testing can be performed in a similar fashion and they will be considered together. The method is based on the proposed L , f and g functions and the detailed descriptions of the approach are provided in Chapter 3. For the purpose of testing H1 and H2, we propose to conduct a driving simulator experiment which will employ predominantly visual interactions (in the rest of this dissertation, we will refer to this study as “Exploring Augmented Reality Navigation Aids”). Based on the previous experience, interactions which are predominantly visual tend to produce localized effects on driving. These effects may be missed by the average-based measures, but are expected to be successfully detected by our cumulative and instance-based measures.

The experiment will involve driving while interacting with three different personal navigation devices. This condition is often found in normal driving, which ensures that the task will not appear artificial to participants. In order to be able to observe how the participants use those navigation devices under the conditions that are close to real life, we intend to implement a realistic (although simulated) city environment. Similar to the previous study [23], we expect to observe a relationship

between glances directed off-road and the changes in driving performance measures, specifically lane position and steering wheel angle. The main idea behind this assumption is that drivers need to continuously control the position of the car in the lane. Hence, if a driver is not looking at the road, the controlling error may accumulate, which would require applying a correction in the car position after the gaze returns to the road. The proposed experiment will involve multiple experimental conditions (or levels of engagement with the secondary task). Thus, it will provide a fairly diversified data corpus. We expect that both the cumulative and the instance-based measures will be able to detect the effects on driving and cognitive load of the three navigation aids through statistically significant peaks in cross-correlation functions.

As indicated in Section 1.3, our ultimate aim is to make this method applicable to any interaction type. While we do not expect that this dissertation will be able to provide such vast generalization, we are taking one additional step in that direction by testing the method under yet another circumstance: that of manual-visual interaction, which expands from the previous visual-only. For this purpose, we propose to test hypotheses H1 and H2 in a driving simulator experiment which examines manual-visual interactions with a popular in-vehicle device: an iPod (we will refer to this study as “Highway Driving and iPod Interactions”). The interactions will involve three distinct levels of difficulty and will be performed while driving on a straight highway road. Similar to the study proposed above, we expect that both cumulative and instance-based measures will detect the effects of interactions with the iPod demonstrated by the significant cross-correlation peaks.

1.4.2 Testing H_{RP} – Ranking the Effects of Secondary Task

Engagements

Hypothesis H_{RP} proposes a way for testing differences between the experimental conditions based on the results obtained using the methods proposed in hypotheses H1 and H2. If the significant differences are confirmed, we can perform ranking of the experimental conditions and compare their relative sizes of the effects.

For the purpose of testing H_{RP} we will employ the same studies proposed for testing hypotheses H1 and H2. Both of these studies involve several difficulty levels, which make it advantageous for testing H_{RP} . We propose the following procedure. First, we will calculate the cumulative and instance-based measures for each type of secondary task engagement. Since our method provides results in the form of functions, as opposed to individual values in case of average-based measures, we cannot directly apply statistical tests to compare the results obtained for different experimental conditions. In other words, we have to characterize our cross-correlation results in a certain way such that they can be acceptable for statistical analysis. We propose two approaches for solving this problem. In the first approach, we can extract the magnitudes of the most prominent cross-correlation peaks for each experimental condition, which tells us how large the influence is at a particular lag. In the second approach we can calculate the areas below the cross-correlation curves for a range of lags, which tells us how large the effect is over a wider interval of time after the occurrence of the event (initiator) of interest (in our case, returning the gaze to the road). Once we have the data extracted using the two approaches, we will conduct comparisons to examine whether statistically significant differences exist in the observed results between different types of task engagements.

Finally, conditional on the existence of the above significant differences, we will rank the effects of different secondary task engagements. This approach will be applied to both studies. The technical details of the approach itself as well as the obtained results are provided in Chapter 3.

1.4.3 Testing H3 – Analyzing Underlying Mechanisms

Hypothesis H3 proposes to analyze a set of variables which contribute to the cumulative and instance-based measures proposed in hypotheses H1 and H2. If proved significant, these variables will provide insight into the underlying mechanisms.

We propose to test this hypothesis through two controlled “reference” experiments that incorporate task-oriented interactions. Both experiments will be performed with an iPod under different driving and interaction conditions. In the first case we will use the same iPod study as in Chapter 3 (“Highway Driving and iPod Interactions”), where driving will be performed on a straight highway road with light traffic. In the second case the simulated environment will resemble a busy, straight city road (we will refer to this study as “City Driving and iPod Interactions”). In other words, the secondary task engagement will be exactly the same between the two studies; the only characteristic that will change is the driving environment. We also plan to include a lead vehicle in both studies, which will provide a uniform driving reference for all participants. The controlled conditions in these experiments will enable us to create multiple regression models in order to determine which of the variables proposed in hypothesis H3 have a significant influence on the cross-correlation results. The complete details of the results of reference experiments and testing H3 will follow in Chapter 4.

1.5 Dissertation Organization

This dissertation is organized as follows. Chapter 2 outlines procedures and measures that researchers commonly employ for characterizing driving performance and cognitive load in the area of human-computer interaction (HCI) in vehicles. Special emphasis has been paid to the measures employed in various studies presented in this dissertation. Chapter 3 gives a detailed description of the cross-correlation method proposed in the introduction. The method is backed up by some specific examples and provides support for hypotheses H1, H2 and H_{RP} . Chapter 4 proposes explanations and a proof-of-concept for the predictive ability of the cross-correlation results, which provides support for hypothesis H3. Concluding remarks and the proposed directions for future research are given in Chapter 5.

Finally, interested reader is encouraged to read through the appendices as well, since they provide more technical and methodological details about the studies presented in this dissertation. Specifically, Appendix A gives a detailed explanation of the data synchronization procedure which was employed in all experiments. Appendix B provides details about the experimental apparatus, description of the NASA-TLX questionnaire and Institutional Review Board approval. Finally, Appendix C provides graphs which were used for testing the assumptions of multiple regression models created in Chapter 4.

CHAPTER 2

BACKGROUND

The problem of drivers getting distracted by in-vehicle devices is certainly not a new one. Ever since the first device with a significant potential for distraction has been introduced in vehicles, such as a car radio in the late 1920s, there have been divided opinions about the effects those devices may have on driving. This notion was summarized well by Nicholas Trott's 1930s article in *The New York Times*: "A grave problem that developed in New Hampshire... now has all the motor-vehicle commissioners of the eastern states in a wax. It's weather radios should be allowed on cars. Some states don't want to permit them at all – say they distract the driver and disturb the peace..."

Cars have changed significantly over the last 100 years. However, most of the changes occurred in the last 10 to 20 years and many are impacting the cabin. Namely, the number of in-vehicle services such as music selection, navigation, live traffic reports and social networking is increasing rapidly. There is a considerable demand for those services, which indicates that the secondary tasks are becoming ever more important to the drivers. This trend is not surprising given that 86% of American citizens spend on average about 25 minutes commuting to work [43]. This suggests that drivers will keep

interacting with in-vehicle devices for two reasons: they have the capability to and it makes their drives a more enjoyable activity. One important factor that fuels this trend is the original equipment manufacturers that deploy various devices in vehicles. However, as argued by Magladry and Bruce [44], the question is not whether we are capable of developing some functionality, but whether we are supposed to. Since new devices utilize various types of interactions, it is necessary to examine their influences on driving even before they find their way into vehicles. As Strayer and Lee suggest [45], if the new technologies are properly designed they can increase safety and enjoyment; however, a poor design can make them deadly. Thus, having reliable tools for estimating driver's distraction is very important and is the topic of this research.

2.1 Driver Distractions

Driving a vehicle is a complex task which requires drivers' full attention (both visual and mental) to be directed to the road ahead. According to Michon [46], driving relies on the processes at three hierarchical levels: strategic level (high level planning of the trip, such as trip goals and desired route), maneuvering level (recognizing current traffic situations and executing maneuvers, such as obstacle avoidance, turning, overtaking, etc.) and control level (low level operation of the vehicle through the available controls, such as steering wheel and throttle). Distractions can occur on any of these levels and they can result in performance decrements at other levels. Nevertheless, drivers very often engage in side activities while driving. As an example, it was estimated that about 9% of drivers were using either a hand-held or a hands-free cell-phone while driving in the US in 2009 at any given daylight moment [6].

In-vehicle activities can be divided in two broad groups: activities supporting the driving task (such as, looking at the speedometer, checking mirrors, using a PND for orientation) and activities supporting drivers' non-driving related needs (such as, talking over the cell phone, checking email, staying in touch with friends over social networking websites). It can be argued that many of the non-driving related needs are "imposed" on drivers by the technological factors and societal norms [45], such as social networking. Whatever the reason behind using those types of devices while driving, they should be carefully analyzed with respect to their ability to distract drivers. The support for this claim comes from a NHTSA study [7] published in 2009 which indicates that 16% of all fatal crashes and 21% of all injury crashes involved driver distraction. Furthermore, during the 100-Car Naturalistic Study [47] where 241 drivers drove 100 instrumented vehicles for the period of 12 to 13 months, over 22% of all crashes and near-crashes were caused by drivers involved in secondary tasks.

Driver distraction can be defined as any activity or process that draws away the driver's attention and disturbs driving control [48]. As such, driver distraction comes in the following forms [49]:

1. *Physical distraction* is the result of physically manipulating an object while driving. This kind of distraction requires removing (at least one) hand from the steering wheel in order to perform the manipulation. Good examples include adjusting a radio [25;50] and operating an MP3 player [18;26].
2. *Visual distraction* prevents a driver from scanning the surrounding environment properly and comes in three forms. The first form includes physical occlusion of the driver's visual field by the obstacles present on the windshield. The second

form includes looking at various objects not directly related to driving, such as in-vehicle infotainment systems (IVIS) [12;24] or navigation devices [20;23;36]. Finally, the third form is usually referred to as “looked, but failed to see” and results from drivers being unable to see a potential hazard even though their visual attention may be directed in the direction of the hazard [51].

3. *Cognitive distraction* is the result of directing the driver’s mind “off-road” to the extent that it negatively influences driving performance [21;52]. This kind of distraction is concerned with the mind being directed to an object of interest and may even be a contributing factor to the “looked, but failed to see” accidents [51].
4. *Auditory distraction* results when drivers focus their attention to different sounds either continuously or occasionally [14]. The most obvious example of auditory distraction is the hands-free cell phone conversation [53-55].

Even though distractions by stimuli external to the vehicle are also occurring (such as advertising [56], road-side events, people) we are focusing here on distractions caused by interactions with various in-vehicle devices while driving. Depending on the user interface design, there exist three basic interaction types: manual, visual and spoken. These three interaction types are orthogonal, which means that they are independent of each other in a sense that they employ different interaction modalities. All types result in cognitive distraction, since in each case it is necessary to mentally process the action; however, only manual interaction produces physical distraction as well. Pure manual interaction requires developing a muscle memory in order to interact with an object of interest. Some representative examples include activating direction lights, wipers or shifting gears. Visual interaction is established through the eye contact with an object of

interest, an example being glancing at a speedometer. Finally, spoken interaction requires verbal contact with a desired object, such as issuing commands to a voice recognition system or listening to navigation directions.

In reality many in-car devices require combinations of the above basic interaction types, such as manual-visual, manual-spoken, manual-visual-spoken. Table 2.1 gives an overview of the interaction combinations used in our preliminary studies introduced in Chapter 1 for the purpose of supporting the definition of the main problem investigated in this dissertation.

Study number	Study name	Interaction type(s) used in the study
1	Interacting with Mobile Radios	manual-visual, spoken (speech production and comprehension)
2	Speech Interface Accuracy and Driving Performance	spoken (speech production and comprehension)
3	The Effects of PNDs on Driving and Visual Attention	manual-visual, visual-spoken (speech comprehension only), spoken (speech comprehension only)
4	Glancing at PNDs Can Affect Driving	visual-spoken (speech comprehension only), spoken (speech comprehension only)

Table 2.1 Interaction types explored in four preliminary studies.

As we can see from Table 2.1, different interaction types and their combinations have been used in these studies. Study 1 investigated interactions with mobile police radios using two alternatives: GUI and SUI. GUI required manual-visual (manually pressing buttons on the radio and observing the LCD display), while SUI required spoken interaction only (both issuing speech commands and comprehending speech recognition engine's responses). Study 2 was focused on spoken interactions only, while studies 3 and 4 explored interactions with various PND alternatives, which can be

divided into: manual-visual (physically manipulating a sheet of paper with written navigation directions - study 3), visual-spoken (navigation information obtained visually by looking at the on-screen directions and verbally by listening to spoken prompts - studies 3 and 4) and spoken (navigation information obtained by listening to spoken-only prompts - studies 3 and 4).

Since in-vehicle interactions often encompass a combination of multiple different distraction types, this makes it more challenging to estimate precisely how difficult a task is. The difficulty of a task is typically not directly observable, because the same task can be more difficult to some individuals than to others. This implies that the overall difficulty of the task depends highly on the interaction between the task and the operator [9]. This is especially emphasized when the operator is instructed to perform both the primary (i.e., driving) and the secondary task (i.e., interaction with an in-vehicle device) simultaneously. Namely, it can be expected that the operator is quite capable of performing each of these tasks individually with high success and relatively low (or at least acceptable) mental demand. However, when both tasks are introduced concurrently the interaction between those may cause an increase in difficulty that the driver is unable to cope with. In other words, it is likely that the overall difficulty of performing two tasks concurrently may be larger compared to the difficulties introduced by each task performed individually. This situation is best described with the concept of high cognitive load (or workload), which may result in deteriorated vehicle control. One of the most famous examples is driving and communicating on a hand-held phone, which can result in driving impairments as profound as those associated with drunk-driving [19]. A very nice summary of the inherent limitations that people have with respect to driving is

given by Rumar [57], who asserts that a driver is an “outdated human with stone-age characteristics and performance who is controlling a fast, heavy machine in an environment packed with unnatural, artificial signs and signals.” Having this in mind it is very important that the original equipment manufacturers focus their efforts in the early stage of device design towards reducing cognitive impairments that may occur as a result of using the device while driving. The next section will provide more details about cognitive load and the methodologies that can be used for detecting it.

2.2 Cognitive Load

Cognitive load is commonly defined as the relationship between mental resources which are required for accomplishing a given task and the resources which are available for that task [10]. Every activity involves a certain amount of cognitive load. The Yerkes-Dodson law [58] provides an empirical relationship between workload (or arousal) and the performance level on a given task. It is an inverted U-shape curve, which increases as the workload increases up to a point, after which starts to decrease. Figure 2.1 illustrates this relationship.

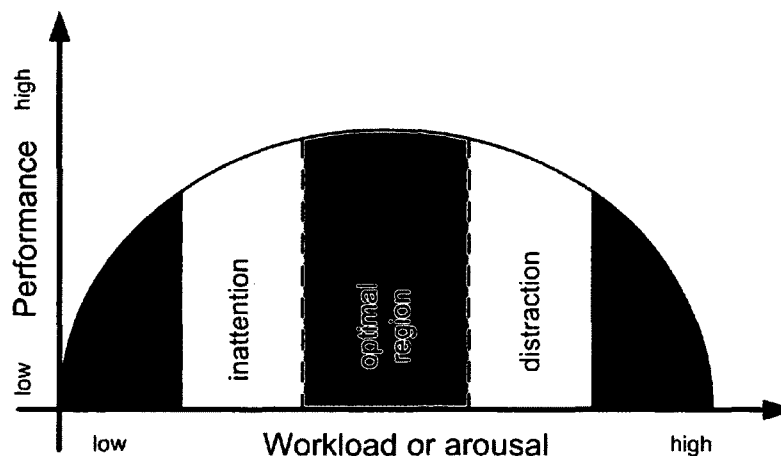


Figure 2.1 Inverted U-shape relationship between performance and workload (arousal).

On the one hand, if someone experiences workload that is too low for a long period of time, it may induce fatigue, boredom and reduced alertness and situation awareness [59], which leads to decreased performance. On the other hand, if the workload is too high (demand exceeds the capacity), one may feel overloaded, which again reduces performance. The relationship between performance and cognitive load is certainly a complex one, since different tasks require different levels of workload for optimal performance.

Cognitive load is a multifaceted, multidimensional problem that is difficult to define [9]. As we had a chance to see in the previous section, it is tightly related to the task difficulty, which can be interpreted as the difference between the expected and the actual performance [9]. In the automotive domain the performance on the primary driving task is of the utmost importance and is assumed to be at its maximum if the driver's attention (both visual and mental) is focused to the road ahead. However, by introducing side tasks the driver is forced to multitask [60;61], which results in divided cognitive resources between driving and side tasks. Since the available resources are limited [62], failures in achieving the expected levels of performance may occur on both sides (driving and side task), which can be attributed to high cognitive load.

Over the years, researchers developed various models that attempt at explaining how the limited cognitive resources are allocated between concurrent tasks. Some early models include the single-bottleneck models [63;64] and single-resource model [65]. However, it became obvious that the time-sharing between tasks is more efficient if they employ different information processing structures than the common ones. This gave rise to the Wickens' multiple-resources model.

According to the multiple resources model [10], one does not possess only a single information processing unit, but rather multiple separate resources that can be utilized simultaneously. There are four dimensions in the multiple resources model which affect the time-sharing performance: processing stages (perception, cognition, responding), perceptual modalities (visual, auditory), visual processing (focal, ambient) and processing codes (spatial, verbal). The model implies that the interference between two tasks will be higher if they require the same level of the same dimension (for example, two visual tasks), than if they require different levels of the same dimension (for example, one visual and one auditory task). The main strength of the multiple resources theory is that it can predict the kinds of tasks that can likely interfere with each other as well as the kinds of tasks that can be performed concurrently.

Regarding cognitive load the theory is the most useful in the overload region (where no residual capacity remains), since it can predict how much the performance will suffer when the overload is reached. It should be noted that the theory has little relevance for characterizing single-task demand, since in that case there are no parallel tasks competing for the same resources. In the automotive domain, however, multiple resources theory fits very well, the reason being the high complexity of the driving task. As such, automobile driving may require resources at multiple levels of processing: perceptual (ambient and focal visual processing, needed to detect lane markers and road signs), cognitive (spatial processing, needed to determine the position of the vehicle in the lane), and response (spatial response, needed to control the steering wheel). Thus, by introducing side tasks it is likely that some of the resources will have to be shared between the two.

We can look at some of our preliminary studies in the light of multiple resources theory, since it may indicate the cases when the performance may be affected as a result of interacting with in-vehicle devices while driving. Table 2.2 gives an overview of the resources used in three example studies. The abbreviations used in the header of the table have the following meaning: V = Visual, A = Auditory, f = Focal, a = Ambient, s = Spatial, v = Verbal, C = Cognitive and R = Response. The check marks indicate whether a task depends on a given resource.

Study	Task	Resources							
		Perception				Cognition		Response	
		Vf	Va	As	Av	Cs	Cv	Rs	Rv
All	Driving only	✓	✓			✓		✓	
Interacting with mobile radios	SUI				✓		✓		✓
	GUI	✓				✓		✓	
Speech Interface Accuracy and Driving Performance	PTT + low accuracy				✓		✓	✓	✓
	PTT + high accuracy				✓		✓	✓	✓
	No PTT + low acc.				✓		✓		✓
	No PTT + high acc.				✓		✓		✓
The Effects of PNDs on Driving and Visual Attention	Standard map-based PND	✓			✓	✓	✓		
	Voice-only PND				✓		✓		
	Paper directions	✓				✓		✓	

Table 2.2 Resource allocation for three example studies.

As mentioned before, operating a vehicle under unencumbered conditions requires multiple resources which are indicated in the first row of Table 2.2. Since all studies involved driving, these resources are common for each study. The following rows indicate the resources used for side tasks in each example study.

The first study (“Interacting with mobile radios”) explored interactions with mobile police radios using two modalities: SUI which used voice commands and GUI

which used embedded hardware controls. In doing so, SUI relied on the following levels of processing: perceptual (auditory-verbal for sensing speech-recognition engine's responses), cognitive (verbal processing for understanding and producing speech commands) and response (verbal response for uttering speech commands). Conversely, GUI relied on a different set of resources: perceptual (focal visual processing for sensing controls on the police radio), cognitive (spatial processing for determining which buttons to press) and response (spatial response for activating desired buttons). If we compare the resources used by the driving and GUI task we can see that there exists a complete overlap between the two. On the other hand, SUI uses entirely different levels of the three processing stages. This implies that the interference is likely between the driving and the GUI task, but not between the driving and the SUI task. Indeed, this assumption was confirmed by both average-based driving performance measures (variances of lane position, steering wheel angle and velocity) and subjective estimates of cognitive load (NASA-TLX questionnaire).

The second study ("Speech Interface Accuracy and Driving Performance") examined the effects of the following SUI characteristics on driving: speech recognition accuracy, PTT button usage and dialog repair. The following resources were common in all cases: perceptual (auditory-verbal for sensing speech-recognition engine's responses), cognitive (verbal processing for understanding and producing voice commands) and response (verbal response for uttering speech commands). Additionally, the conditions which involved using the PTT button (PTT + low accuracy and PTT + high accuracy) also relied on spatial response for manually pressing the PTT. However, we can say that this action mostly relied on muscle memory, since the participants were trained to operate

the PTT without the need to look at it. If we compare the resources required by these secondary tasks with the ones required for driving, we can see that an overlap exists in spatial response for the conditions which involved using the PTT button. This suggests that the likelihood of interference with driving is higher in those conditions than in the “no-PTT” conditions. Our results support this assertion for lane position variance, which was significantly higher when the PTT button was used and speech recognition accuracy was low. We also found that the steering wheel angle variance was significantly affected by the recognition accuracy whether the PTT button was used or not.

The third study (“The Effects of PNDs on Driving and Visual Attention”) explored three in-car navigation alternatives: standard map-based PND, voice-only PND and paper directions. Standard map-based PND (SPND) required the following resources: perceptual (focal visual processing for observing the map and auditory-verbal for detecting spoken directions) and cognitive (spatial and verbal processing for interpreting the position of the vehicle on the map and understanding spoken directions). Voice-only PND relied on the following resources: perceptual (auditory-verbal for detecting spoken directions) and cognitive (verbal processing for understanding spoken directions). Finally, paper directions required: perceptual (focal visual processing for observing written directions), cognitive (spatial processing for reading the directions) and response (spatial response for handing the sheet of paper). If we compare the resources used by the driving task with each of the above three tasks we can see that the largest overlap exists for paper directions and SPND. Since there is no overlap between driving and voice-only PND, we can consider it as a “baseline” condition for comparisons. Thus, we would expect the largest interference with the primary task in the case of paper directions, which

was detected using both the average-based driving performance measures and visual attention. The interference with driving could also be expected in case of standard map-based PND, which was supported by visual attention but not driving performance measures. Given the significant impact on visual attention ($p < 0.05$) when the SPND was used (PDT on the road ahead was 88% vs. 92% for voice-only PND) and the fact that driving is a predominantly visual activity, our conclusion in the introduction was that the average-based measures may not be sensitive enough to detect influences of individual glances towards the map-based PND.

Even though the influences of gazing away from the road are sometimes not detected using averages (as was the case with the SPND in the study above and multiple studies presented in Chapter 1), it does not mean that they do not exist or should be ignored. This is especially important with well-designed in-vehicle devices that may even encourage drivers to interact with them more frequently while driving. As Lee and Strayer point out [45], this can lead to a usability paradox, which occurs when improved ease of use makes each individual interaction less distracting, but as a result of more frequent use the overall risk of using it increases. This problem can also be looked at from the perspective of the “Swiss cheese” model of incident occurrence [66], which postulates that accidents occur when all necessary adverse conditions line up thus allowing a negative consequence to occur. We argue that interacting with in-vehicle devices fits this analogy fairly well and can be explained as follows.

Let us assume there are three layers in the model: glances directed away from the road (i.e., towards an LCD screen), changes in driving performance (i.e., swerving in the lane) and the presence of a hazardous object (i.e., a pedestrian or another vehicle).

The third layer can also represent the driving environment in general, such as good vs. inclement weather, day vs. night, etc. If a driver is paying sufficient attention to the forward road, controls the vehicle well and no hazardous objects are present on the road, the holes will not appear in the layers (at least not in the first two layers). The appearance of “holes” is an indication of unfavorable conditions.

Regarding the first layer, the more often a driver looks away from the road, the larger the number of holes. Similarly, the longer the individual glances are, the larger the corresponding holes. This assumption matches the observation from the literature, which states that the aggregate risk of using a particular in-vehicle device is equal to its exposure, that is, the product of the duration of each use and the frequency of use [67].

The holes in the second layer do not have to be aligned with the holes in the first layer (at least not all the time), which indicates the fact that not every glance directed away from the road will necessarily instigate worse driving performance. Additionally, glances that do result in worse driving do not necessarily have the same size of the effect, thus differently affecting the sizes of the holes.

Finally, the holes in the third layer indicate how often (number of holes) and for how long (sizes of holes) the hazardous objects are present on the road. This layer is directly affected by the driving environment: it is more likely that the hazardous situations will occur on the busy city streets during a rush hour than on a free-flowing highway. Similarly, driving is much easier on a straight road, with no traffic and under good weather conditions, than under heavy traffic and torrential downpour [68].

We can also argue that the probabilities of holes appearing in these three layers are going down with each successive layer: glances directed off-road are very likely to

happen, but only a portion of them will impact driving. Similarly, hazardous events occur rarely, so the probability of holes appearing in the third layer can be expected to be the lowest. Therefore, we can conclude that the overall probability of an accident occurrence is fairly low, but still higher than zero. The question is whether we can make this probability even smaller?

The third layer predominantly depends on chance, since the appearance of hazardous objects or the environmental conditions cannot be controlled by the driver. However, the first two layers can. The situation in which the holes in the first two layers align can be termed as a “near-hit,” since the driving performance is affected by the glances directed off-road, but ultimately no collision occurs. The probability of this situation is certainly higher than the overall accident probability and should be made as low as possible. This is exactly the reason why it is necessary to have reliable tools which would detect negative impacts on driving within the first two layers. Multiple resources theory is certainly one useful qualitative tool which can detect interferences that may result in high cognitive load conditions. However, sensitive empirical tools are also necessary which would support its predictions.

In general, cognitive load also depends on the context where the task is performed. In other words, the load at each resource depends on the complexity of the driving environment and is likely to be different from study to study. To cite an example by Wickens [10], visual/spatial resource demands are likely to be relatively high on dimly illuminated roads. However, they may be even higher if the road is curvy and the travelling speed is high.

The above effect of the driving environment may be visible from two studies proposed in Chapter 1 for the purpose of testing multiple hypotheses: “Highway Driving and iPod Interactions” and “City Driving and iPod Interactions.” In both cases the main task is driving, while the secondary task includes interactions with an iPod of varying levels of difficulty. The secondary tasks in both studies rely on the same resources: perceptual (focal visual processing for detecting buttons and scanning the LCD screen on the iPod), cognitive (spatial processing for determining which buttons to press and understanding the information presented on the LCD screen) and response (spatial response for pressing the buttons). If we compare these resources with the ones required by the driving task (see the first row in Table 2.2) we can conclude that the overlap is significant and that the interference (and thus increased cognitive load) with driving is likely. Based on this we expect that through these studies our cross-correlation method will be able to accomplish the following: a) detect both cumulative (supporting H1) and instance-based (supporting H2) changes in cognitive load resulting from the interference between the driving and secondary task, and b) allow ranking of the levels of secondary task engagement (supporting H_{RP}). Furthermore, if the interference is also detected by the standard average-based driving performance measures, this will provide support for construct validity of our method. However, we also expect that the detected impacts will be different between the studies, because they employ different environment conditions: straight highway with light traffic in the one case and busy city road in the other. Thus, it is likely that the resource allocation (and thus cognitive load) may be higher in the latter study. As pointed out by Zhang et al. [30], the amount of distraction that the drivers are willing to sustain may be smaller under difficult conditions and larger under easy

conditions. Nevertheless, drivers very often allow the performance of the primary driving task to degrade [68].

Acknowledging that cognitive load is such a complex concept, it is unlikely that any single measure would be good enough for its characterization. Thus, researchers utilize a large number of measures that can be classified in three main categories [11]: performance-based (which can be divided into primary-task and secondary-task measures), physiological and subjective. The following sections will give a brief overview of these categories in the context of driving research.

2.2.1 Performance-based Measures

Performance-based measures assess workload by analyzing how well the operator performs a given task. These types of measures are very easy to comprehend, since they directly reflect the results of the operator's efforts. There are two variants of the performance-based measures: primary and secondary task.

Primary task measures estimate an operator's capability to perform the actual task of interest. As the cognitive load increases, more resources are utilized which may eventually lead to a performance decrease. One disadvantage of this type of measure is that it is insensitive to workload changes in the situations where the operator can provide additional effort (has some spare cognitive capacity) to maintain the desired level of performance. Nevertheless, it is often used, especially when it is desired to distinguish different levels of cognitive load when the performance has already been affected (e.g., driving on a curvy road at low and high speeds) or to discriminate conditions of non-overload and overload (e.g., driving on an empty road and in traffic jam).

Secondary task measures require the operator to perform two tasks in parallel: primary task and side task. Primary task is of the utmost importance and the operator's performance on that task is continuously monitored. The side task is performed concurrently with the primary task and can be used to probe the spare cognitive capacity remaining after the primary task. This way, the performance while executing the side task is a proxy for measuring the spare cognitive capacity. Since driving itself is often within the cognitive limits of the operator, by introducing side tasks, an overload condition may occur. This approach is sometimes used in driving research. For example, Reimer et al. [69] used a delayed digit recall secondary task (n-back) while driving to evaluate gradual changes in cognitive load as detected by physiological measures. Similarly, Harbluk et al. [21] used single and double digit addition problems as the cognitive task. However, often the goal of a study is not to probe the spare capacity with a secondary task, but rather to investigate the secondary task as the addition to the primary task of driving. In that sense, the secondary task may be considered as another "primary task" of interest. The approach used in this case is referred to as the embedded secondary task.

According to Eggemeier and Wilson [70], an embedded secondary task is a function conducted by the operator concurrently with the primary task, but is distinct from the primary task which is being assessed. The advantages of this approach are that both tasks constitute normal operator behavior, do not appear artificial to operators and have high operator acceptance. The embedded secondary task approach is used exclusively in the studies presented in this dissertation, since each study is concerned with a particular in-vehicle device whose impact on driving is of interest. For example, in the police radio study ("Interacting with Mobile Radios"), we investigated changes in

primary task performance (driving) resulting from interactions with the radios (embedded secondary task). Similarly, in order to test our hypotheses specified in the introduction, we will analyze the effects of multiple personal navigation devices (“Exploring Augmented Reality Navigation Aids”) as well as iPod interactions on driving (“Highway Driving and iPod Interactions” and “City Driving and iPod Interactions”). All of these secondary tasks are commonly performed in vehicles, thus ensuring drivers’ acceptance and similarity to real life driving.

In the context of driving research, primary-task measures are referred to as driving performance measures and they typically include: lane position, steering wheel angle, velocity, lateral velocity, following distance, headway time, time to collision, number of lane crossings, number of collisions and many others. Most of these variables are continuous in their nature and are typically transformed in some way in order to obtain more descriptive metrics that would be suitable for follow-up statistical analysis. The most common transformations are the mean [12;14;16;19;20;23;35;71] and variance [16;22;23;25;35] or standard deviation [12;14;15;18;19;35;71;72] of a desired driving variable. Usually, in case of variances or standard deviations of driving performance measures, a higher numerical value in one experimental condition in comparison to others indicates worse driving. One good example is the study where we explored the influence of speech recognition engine’s accuracy on driving performance [22] (“Speech Interface Accuracy and Driving Performance”), where the low recognition accuracy condition was associated with higher variances of the steering wheel angle compared to the high accuracy condition. This suggests that in the low accuracy condition the participants expended more effort on steering in order to keep the vehicle in the lane.

The above mentioned transformations are applied to each experimental condition, either as a whole or with some appropriate segmentation (for example, the beginning of a segment can be each time a driver uttered a command to an in-vehicle interface). This way each condition (or each segment) is characterized with a single number that can be used for comparison with other conditions using various statistical methods, such as ANOVA [12;14-16;22;23;25;26;71-73]. However, as explained in Section 1.1.2 of the introduction, the consequence of applying average-based transformations is that the important effects of in-vehicle interactions on driving may not be detected in the averages. This of course does not mean that the average-based measures are not useful. In fact, they are used throughout this dissertation as can be seen in Table 2.3. The top row shows the specific measures employed in each study. Studies 1 through 4 were used in Chapter 1 to support the definition of the main problem. Studies 5 through 7 will be used in the following chapters for testing all of our hypotheses. By comparing the results obtained using average-based measures with the ones obtained using our proposed cross-correlation method we will be able to draw conclusions about their sensitivities to changes in cognitive load.

Since in our proposed cross-correlation method we use glances directed away from the road to indicate where changes in driving performance may occur, we can categorize the cross-correlation results under the performance measures as well. Section 2.4 gives an overview of multiple scientific areas where the cross-correlation function has been applied successfully.

Study number	Study name	Variance of lane position	Variance of steering wheel angle	Variance of velocity	Mean of velocity	Number of collisions
1	Interacting with Mobile Radios	✓	✓	✓		
2	Speech Interface Accuracy and Driving Performance	✓	✓	✓		
3	The Effects of PNDs on Driving and Visual Attention	✓	✓	✓	✓	
4	Glancing at PNDs Can Affect Driving	✓	✓	✓	✓	✓
5	Exploring Augmented Reality Navigation Aids	✓	✓	✓	✓	✓
6	Highway Driving and iPod Interactions	✓	✓	✓	✓	✓
7	City Driving and iPod Interactions	✓	✓	✓	✓	✓

Table 2.3 Average-based driving performance measures employed in studies presented in this dissertation.

2.2.2 Physiological Measures

Physiological measures enable workload assessment based on the biological processes, such as heart rate, respiration, pupil dilation, etc. Some of these measures appear sensitive to global changes in workload levels (such as pupil dilation), while some appear diagnostic to a specific resource usage (such as event-related brain potentials). Most physiological measures are controlled by the autonomic nervous system (ANS). This means that they are not under voluntary control, which makes them fairly objective.

Some of the more popular physiological measures include: heart rate [31-33], skin conductance [31;32], transient cortical evoked response [9;11], pupillary response [74;75], heart rate variability [76;77], and so on.

It has been shown in the research literature that both heart rate (HR) and skin conductance (SC) increase as the cognitive load increases [31;32;69;78]. As part of testing our hypotheses, we propose to collect those variables in our final study of iPod interactions while driving (“City Driving and iPod Interactions” – see Chapter 4). The reason for including those measures is to demonstrate that cross-correlation results indicate changes in cognitive load in the same direction as the physiological measures. This provides another source of support for construct validity of our method and also goes along well with Wickens’ assertion [10] about avoiding circular arguments, as discussed in Chapter 1.

Heart rate (HR) is obtained from the electrocardiogram (ECG), which represents the electrical activity of the heart muscle. It is obtained by counting the number of R-impulses (prominent, periodic changes) in the raw ECG signal and is expressed as the number of beats per minute. Inter-beat interval (IBI) is inversely related to HR and can also be used. It is measured as the time interval between consecutive R-impulses. Heart activity is controlled by the autonomic nervous system (ANS), such that the sympathetic branch increases the heart rate, while the parasympathetic branch decreases the heart rate.

Electro-dermal response (EDR) [79] is also controlled by the ANS and represents changes in electrical properties of the skin (eccrine sweat gland activity), which are caused by environmental and psychological states of an individual. Even

though skin resistance can be measured as well, some of its properties make it less desirable [80]: it is strongly influenced by the features which are not relevant to the physiological activity, it is far less linearly related to the activity of sweat glands and its measures are less normally distributed than the measures of skin conductance. Thus, skin conductance is the preferred option when analyzing EDR and it is measured in micro Siemens [μ S]. There are two types of EDRs: tonic and phasic [79]. Tonic response is the “baseline” level without any stimulating events. Phasic responses occur when stimulating events take place and are characterized by rapid peaking (with some latency) in skin conductance followed by returns to the tonic level. However, phasic responses often occur without any specific stimuli and are thus called non-specific EDRs.

Various measures can be extracted from the raw HR and SC signals: heart rate variability (HRV) in case of heart rate and latency, rise times, recovery times and frequency of EDRs per minute in case of skin conductance. We decided to apply the same approach as Mehler et al. [32], who calculated average values of heart rate and skin conductance for multiple levels of secondary task difficulty. Their results indicated incremental increases in both variables with the increase of cognitive load introduced by a delayed digit recall secondary task (n-back). Thus, in our experiment (“City Driving and iPod Interactions”) we also calculate the average values of both heart rate and skin conductance signals for each experimental condition. We then perform statistical analyses to determine whether there exist significant differences in experienced cognitive load between those experimental conditions (see Chapter 4). In order to demonstrate support for construct validity, we also compare those results to the ones obtained by our cross-correlation method. The main advantage of these measures is that they are simple to

implement (relatively simple and unobtrusive instrumentation) and interpret. However, since these measures are also continuous in their nature, the averaging can produce the same problem as before – localized changes may not be successfully detected.

Another group of measures which can be classified under the physiological category is visual attention. Visual attention describes the behavior of a driver's gaze while driving and can be characterized with various measures, such as the duration and/or number of glances [24;34;81], duration and/or number of fixations [27;29;82], eyes-off-the-road time [12;28;30], gaze location [21;82].

According to SAE J2396 and ISO 15007 standards [83], a glance can be defined as a series of fixations directed at a target area until the eyes are moved to a new area. The same standards define a fixation as the alignment of the eyes, such that for a certain period of time the image of the fixated object falls on the fovea. In other words, fixations are limited in both temporal and spatial direction, since they are directed to approximately the same location longer than some predefined time interval [29] (for example, within 1° of visual angle and longer than 0.5 sec). In this dissertation we are concerned with all glances directed off-road, which reflects our expectation that in general they negatively affect driving performance. Therefore, we use all off-road glances in order to implement our cross-correlation method.

SAE J2396 defines glance duration as the amount of time from the moment when the gaze moves toward a desired target to the moment it moves away from it. This information can be used to obtain the total eyes-off-the-road time, which shows the amount of time a driver spends looking away from the road. Equivalently, eyes-off-the-road time can be transformed to eyes-on-the-road time and expressed as a percentage of

the total experiment time. This metric is then called the percent dwell time on the forward road (PDT), which shows on average the percentage of time a driver spends looking at the road ahead [15]. PDT has been used extensively in almost all of our studies (see Table 2.4 below).

Finally, we use the information about the duration and number of off-road glances to obtain a finer picture about the way different experimental conditions influence drivers' visual attention, besides just the overall percent of time spent looking at the road expressed through PDT.

Table 2.4 gives an overview of the physiological measures employed throughout this dissertation.

Study number	Study name	Glance duration	Number of glances	PDT	Durations of fixations	HR	SC
1	Interacting with Mobile Radios						
2	Speech Interface Accuracy and Driving Performance						
3	The Effects of PNDs on Driving and Visual Attention	✓		✓	✓		
4	Glancing at PNDs Can Affect Driving			✓	✓		
5	Exploring Augmented Reality Navigation Aids	✓	✓	✓			
6	Highway Driving and iPod Interactions	✓	✓	✓			
7	City Driving and iPod Interactions	✓	✓	✓		✓	✓

Table 2.4 Physiological measures employed in studies presented in this dissertation.

The header of the table shows the names of the specific measures, while the rest of the rows indicate the actual measures used in each study. The statistical analyses were performed on average values (means) of each physiological measure. The first two rows in the table are empty, because an eye-tracker and a physiological monitor were not available for those studies. The last three rows indicate physiological measures that we propose to collect in the studies intended for testing our hypotheses.

2.2.3 Subjective Measures

Subjective measures have been used very frequently in the research literature to assess operators' workload. Some of the reasons for their popularity include their sensitivity and ease of implementation.

Workload related research has been especially active in the area of pilot workload, which resulted in various rating scales being developed over the years, such as the Cooper-Harper scale [84], Subjective Workload Assessment Technique (SWAT) [85] and NASA Task Load Index (NASA-TLX) [86]. Some of the above rating scales found their way into the automotive environment, such as the NASA-TLX [21;24;30;39;72;87-89]. Other scales, intended to specifically address the automotive context, are available as well, such as Driver Activity Load Index (DALI) [90;91] (which was derived from NASA-TLX), Behavioral Markers of Driver Mental workload (BMDMW) [92] and PSA-Task Load Index (PSA-TLX) [92].

NASA-TLX is used very frequently in the research literature. It is a multidimensional assessment tool, which consists of six scales: mental demand, physical demand, temporal demand, performance, effort and frustration. Each scale is divided in

20 equal intervals anchored by bipolar descriptors (i.e., Very Low/Very High). After the participants provide ratings on each of the scales, they are asked to perform all possible 2-way comparisons (15 in total) of the six scales. This way they compare which of the two dimensions contributed more to the overall feeling of workload. The results of the comparisons are used for calculating the weighing factors, which are then used to obtain the overall estimate of cognitive load. Given its popularity among other researchers, the majority of the studies presented in this dissertation use the NASA-TLX scale. Appendix B gives the descriptions of the six scales given to the participants as well as the NASA-TLX questionnaire itself. It is administered in each of the three studies presented in Chapters 3 and 4 (“Exploring Augmented Reality Navigation Aids”, “Highway Driving and iPod Interactions” and “City Driving and iPod Interactions”) for the purpose of demonstrating construct validity of the cross-correlation method proposed in the introduction. Namely, we compare whether the results obtained using the cross-correlation method support the same trends observed using the NASA-TLX questionnaire. The positive relationship between the two provides support that the cross-correlation results indicate changes in cognitive load.

Often researchers use Likert scales in order to obtain an answer to a particular question [16;21;24;28;35;82]. Likert scales consist of a number (typically 5 or 7) of ordered choices that the participants are supposed to select from when providing their opinion about the given question. For instance, in a study presented in Chapter 3 (“Exploring Augmented Reality Navigation Aids”) we intend to rate participants’ agreement with two preferential statements pertaining to the experimental conditions. The corresponding Likert scales will consist of 5 options: 1 – highly agree, 2 – agree, 3 –

undecided, 4 – disagree and 5 – highly disagree. The data will then be aggregated from all participants and analyzed as a whole. When analyzing the data that originate from the Likert scales, a word of caution is necessary. Since the ratings are not continuous, but rather ordinal, summarizing the central tendency from a Likert scale data should not be done using averages, but rather using medians or modes [93]. Similarly, non-parametric tests should be preferred to parametric tests for statistical inferences, such as the chi-square test or Kruskal-Wallis test [93].

Besides rating scales, self-report measures, such as interviews and post-experiment questionnaires [21;28;94;95] also fall within the category of subjective measures. They are usually less formal than the rating scales, however, the main advantage of questionnaires is that the participants are given an open ended question which they can read and provide an answer without any interference from the experimenter. This way, important insights can be obtained about the specific factors that affected participants' cognitive load and their experiences in general, based on which educated conclusions can be made. Post-experiment questionnaires are also employed in the majority of studies in this dissertation.

Finally, another self-report measure, which comes from the field of psychology, is the Experience Sampling Method (ESM) [96]. As opposed to surveys and interviews, which are recall-based techniques (the experiences are reported after the fact), ESM does not require recalling the experiences from the memory. Rather, brief questionnaires are administered several times (randomly, periodically or when events of interest happen) over the duration of the study in order to capture the participants' behaviors, moods, feelings, etc. as they occur in real-time. The experimenters are not

present while the ESM is being administered. Like other questionnaires, ESM can be used for obtaining both structured (quantitative) and non-structured (qualitative) data. Even though this type of questionnaire may not be always applicable while driving, one example which uses the same underlying logic (although the authors do not specifically state that they are using ESM) is in the 100-car naturalistic study [4]. Namely, the authors installed an “incident” pushbutton below the rear-view mirror that the participants could press whenever an unusual event occurred in the driving environment. In our studies this particular measure was not practical, since it would alter drivers’ normal behavior and possibly introduce local changes in driving performance which could be confounded with the actual events of interest, such as glances directed off road.

Subjective measures can be quite effective, since the operators have the opportunity to directly express their opinion about the difficulty of the desired task. On the other hand, they are usually done with respect to the experiment as a whole, thus making them less suitable for detection of rapid cognitive load changes (except possibly ESM). Furthermore, the fact that these measures are subjective makes them more difficult for comparison between different experiments. This is corroborated by the discussion presented in Section 2.1, which states that the task difficulty highly depends on the interaction between the task and the operator.

Table 2.5 gives an overview of the subjective measures used in preliminary studies (1 through 4) as well as the studies (5 through 7) proposed for testing our hypotheses. As before, the heading shows the names of the measures, while individual rows indicate the specific measures employed in each study.

Study number	Study name	NASA-TLX	Likert scale questionnaire	Interview / Post-experiment questionnaire
1	Interacting with Mobile Radios	✓		✓
2	Speech Interface Accuracy and Driving Performance			✓
3	The Effects of PNDs on Driving and Visual Attention		✓	
4	Glancing at PNDs Can Affect Driving		✓	✓
5	Exploring Augmented Reality Navigation Aids	✓	✓	✓
6	Highway Driving and iPod Interactions	✓	✓	✓
7	City Driving and iPod Interactions	✓	✓	✓

Table 2.5 Subjective measures used in studies presented in this dissertation.

2.2.4 Criteria for the Selection of Workload Measures

Each workload measure can be described using five criteria: sensitivity, diagnosticity, intrusiveness, implementation requirements and operator acceptance. These criteria should be considered in the selection of the appropriate procedure for a desired application. O'Donnell and Eggemeier [11] give an excellent overview of the above criteria, which will be summarized briefly in the following paragraphs.

Sensitivity describes the potential of a measure to identify changes in cognitive load caused by a task of interest. Based on the task characteristics, a measure with the appropriate sensitivity should be chosen. If the goal of the analysis is to determine whether the task causes cognitive overloads which degrade performance, then a primary-task measure should suffice. However, if the goal is to establish whether there is a potential for cognitive overload, some of the more sensitive measures should be considered, such as subjective, physiological or secondary-task. The reason behind this is

that the operators may be able to invest more effort in order to keep their task performance at the desired level. Even though this comes at a price of increased workload, it cannot be detected using the primary-task measures.

The sensitivity issue of average-based driving performance measures is exactly the main problem we are addressing in this dissertation. We argue that our cross-correlation method proposed in the introduction will provide higher sensitivity to changes in primary-task measures caused by cognitive load.

Diagnosticity comes from the multiple-resources theory [10] and it determines the capability of a measure to distinguish which of the available resources is being used by the task of interest. For example, pupil diameter has the potential to assess overall workload on the processing system. In other words, this type of measure does not have high enough diagnosticity necessary for distinguishing a particular resource affected by the task. Conversely, the event-related brain potentials appear to be highly diagnostic to some particular resource usage. Therefore, we can say that physiological measures can either have high or low diagnosticity. Subjective measures typically have low diagnosticity as a result of the operators' inability to discriminate between different resources. Similarly, primary-task measures exhibit low diagnosticity, since it is usually not obvious which particular resource caused decrements in task performance. On the other hand, secondary-task measures are usually highly diagnostic, since they can be designed to probe the spare cognitive capacity on the specific resources. The required level of diagnosticity depends highly on the general objectives of the analysis. If the goal is to estimate the overall workload experienced by the operator, then a less diagnostic

measure can be used. On the other hand, if the goal is to pinpoint the specific resource which is being heavily loaded, a more diagnostic measure should be applied.

Our cross-correlation method is initiator-based, which means that it uses instances of secondary task engagement as reference points for calculations. If the individual interactions use only a single modality (such as visual interaction in “Exploring Augmented Reality Navigation Aids” study), this has the potential to provide fairly high diagnosticity. Namely, we proposed (hypotheses H1 and H2) to use glances directed away from the road as individual instances of secondary task engagement. The reason for this is that we expect that the changes in driving performance measures will be affected by the cognitive load caused by sharing visual resources between the driving and the secondary task. On the other hand, if the interaction is multimodal (such as manual-visual interactions in “Highway Driving and iPod Interactions” and “City Driving and iPod Interactions” studies), multiple resources are used while engaging in the secondary task. In this case we expect that our method will provide less diagnosticity.

Primary task intrusion is the amount of primary task performance degradation attributed to the workload measure itself. Depending on the experimental condition to which the workload measure is applied, different levels of intrusion may be tolerated (e.g., field study vs. simulation). Nevertheless, extreme levels of intrusion should be avoided (or at least minimized), since they may lead to difficulties in the interpretation of the results. By the definition, primary-task measures are not intrusive. Subjective and physiological measures are in general the least intrusive, since often they do not require any additional activity by the operator while performing the primary task. In contrast,

secondary-task measures usually induce significant intrusion, especially if they appear very artificial compared to the primary task.

Since our cross-correlation method uses visual attention and driving performance data, we can say that it is not intrusive.

Implementation requirements specify the complexity of the measurement procedure, such as the required equipment and supporting software. An appropriate measurement technique should be selected based on these requirements and practical constraints. Subjective measures are in general the simplest to implement, since they are often performed after the conclusion of the experiment and require very simple tools. Primary-task measures are fairly simple to implement as well. On the other hand, physiological and secondary-task measures usually require significant instrumentation, software support, operator training or equipment calibration.

The implementation requirements of our cross-correlation method are somewhat higher on the software side compared to average-based driving performance measures. However, the algorithm can be implemented once and reused in many different studies.

Operator acceptance is defined as the participant's recognition of the usefulness of a measurement technique. Attention should be paid to this criterion especially when the participants represent proficient operators of the desired system. Care should be taken to make the measurement technique less artificial and intrusive, since it increases the participants' acceptance.

In all of our experiments we use the embedded secondary task approach, where the participants interact with interfaces commonly found in vehicles. Additionally, since our cross-correlation results are obtained in post-processing, we can expect that the operator acceptance is high.

2.3 Experimental Method

So far we had a chance to observe how distractions in vehicles occur, how they can result in increased cognitive load and how those effects can be detected using various types of measures. However, it is also of interest to examine the typical experimental methodologies that the researchers employ when analyzing in-vehicle interactions.

2.3.1 Experimental Apparatus

Depending on the research capacities, studies are done on personal computers [20;28;73], in driving simulators [12-20;22;23;25-27;35;37;55;72;97-100] or in real cars [4;31;37;101-103]. Since all of the studies presented in this dissertation were performed in a driving simulator, we will focus our attention on driving simulator studies.

Driving simulator studies are very popular because they do not involve any risks to participants, are repeatable, easily customizable, and provide various data which would require complicated instrumentation if desired to be collected in real vehicles. Even though driving simulators do not provide the same level of realism as real driving, they still have a fairly high validity with the results mostly matching the ones obtained in on-road studies [37;104-109]. Lew et al. [108] performed driving simulator experiments with participants suffering from a traumatic brain injury. Their results show that the driving simulator performance measures were good predictors of future driving

performance in real-life when participants have regained some of their abilities lost due to the injury. Wang et al. [107] compared three manual address entry methods in an on-road study and in a medium fidelity, fixed-base driving simulator. Their results indicated that the visual attention and task measures matched very closely between the two environments. Reed and Green [37] used a telephone dialing task to compare driving performance measures between a low-cost driving simulator and on-road driving. They found that lane-keeping performance was less precise in the simulator than on-road. However, speed control was comparable. The same trends were observed with respect to telephone operation: higher variation of lane position and speed were observed while dialing the phone both in the simulator and on-road. The overall conclusion was that their simulator provided a good absolute validity for speed control and good relative validity for driving precision. Driving simulator studies can also help in understanding of human perception and self-motion, which is especially important at speeds and accelerations higher than with natural locomotion [109].

The simulator used in the studies presented in this dissertation is a high-fidelity driving simulator [110]. It provides a very immersive environment with a full car cabin, 180° field of view screen, realistic sounds and vibrations and a motion base for simulating braking and acceleration (see Appendix B for a detailed description of the simulator's capabilities). This kind of simulator has been used widely by the researchers and practitioners in the area of driving research [26;50;111-113]. For example, Slick et al. [106] demonstrated that this particular type of driving simulator can be used as a substitute for naturalistic on-road experiments. They conducted multiple high-risk training scenarios, such as the right/left turn at a stop sign or right/left turn at a traffic

light, using two alternatives: driving simulator and a real car. Their results indicate no significant differences between participants who were trained using either of the two alternatives. These findings are very important, since they offer evidence about the validity of the conclusions drawn from this type of driving simulator.

Driving simulators can sometimes create adverse effects on participants known as simulator sickness, which is usually manifested through headaches, blurred vision, eye strain, nausea, and so on. Mourant and Thattacherry suggest [114] that the vehicle velocity may be an important factor, with higher velocities introducing more sickness. Burnett et al. [115] indicate that the simulator sickness may be mediated by using real car cabins in driving simulators (as is the case with our driving simulator), which can also help with the validity of the results.

From our experimental experience, simulator sickness typically occurs in highly demanding environments which involve frequent 90° turns, such as in the city environment. In agreement with Mourant and Thattacherry [114] we discovered that the likelihood of simulator sickness increases with the speed at which the turn is negotiated. This can be explained by the very fast movement in the peripheral vision which overwhelms the visual experience, however, without the presence of the corresponding forces on the participant's body. Even though our driving simulator possesses a motion base with one degree of freedom (simulating longitudinal movement while braking and accelerating) it does not help with turns. This disconnect between what the participant sees and feels results in simulator sickness. We found no suitable questions about participants' everyday behavior (such as playing sports, video games, riding on roller-coasters and seasickness) that could be asked during the recruitment phase in order to

determine whether a particular person would be susceptible to simulator sickness. However, increased sweating proved to be one physiological characteristic which is a very good precursor for simulator sickness. In order to prevent simulator sickness from occurring during the experiment, we gave each participant a training session in order to get accustomed to the driving simulator. During this session we monitored participants' behavior through the eye-tracker cameras mounted on the dashboard and periodically asked them questions about their condition, such as "Are you feeling warm or sweaty?", "Are you feeling dizzy?", "Are you experiencing a headache?". The participants who successfully finished the training session were then allowed to participate in a study.

2.3.2 Experimental Approach

When designing test drives in driving simulators the researchers typically use the following approaches: unconstrained driving [13;23;27;99], driving with a predetermined speed and position in the lane [15;20;25;72] or following a lead vehicle [12;14;18;19;22;116]. Unconstrained driving is the closest to real life, since participants are instructed to drive as they normally would while obeying all traffic laws; however, in this case the driver has the liberty of changing his/her behavior without constraints, which introduces additional variables that cannot be easily accounted for, such as changing lanes and velocity. Instructing drivers to maintain a constant speed and to remain in a particular lane during the experimental run does not result in realistic driving. Nevertheless, it facilitates the detection of a secondary task influence by analyzing the variables that the driver is supposed to keep constant. A similar approach is used with the lead vehicle option, where a driver is instructed to keep constant distance (gap) behind the vehicle in front.

Properly designed experiments provide motivation for avoiding accidents and maintaining the same kind of driving behavior as they would in the real setting. Thus, we strive to make our experiments less artificial and as close to real driving as possible. Table 2.6 outlines the driving types used in the studies presented in this dissertation.

Study number	Study name	Driving type
1	Interacting with Mobile Radios	lead vehicle
2	Speech Interface Accuracy and Driving Performance	lead vehicle
3	The Effects of PNDs on Driving and Visual Attention	unconstrained driving
4	Glancing at PNDs Can Affect Driving	unconstrained driving
5	Exploring Augmented Reality Navigation Aids	unconstrained driving
6	Highway Driving and iPod Interactions	lead vehicle
7	City Driving and iPod Interactions	lead vehicle

Table 2.6 Driving types employed in studies presented in this dissertation.

As we can see, in the preliminary studies (1 through 4) we used two approaches: following a lead vehicle and unconstrained driving. Following a lead vehicle is fairly close to real life, since it happens often that friends travel separately in individual vehicles and the leader knows the way. In the unconstrained driving approach, the drivers were instructed to drive as they normally would, follow the speed limits and obey all traffic rules. To make the driving task even closer to real life in study 4 we introduced realistic traffic, pedestrians and unexpected events (pedestrians jaywalking, cars braking, etc.) which are all very common in a busy city environment.

We decided to use the same approaches in the studies proposed for testing our hypotheses (studies 5 through 7 in Table 2.6). Study 5 will be testing personal navigation devices. Since the navigation directions are the most useful in city driving, we will implement a realistic city environment with unconstrained driving in this study. Studies 6

and 7 will be testing interactions with an iPod while driving. For this purpose we will use a lead vehicle approach - once on a straight highway road and once on a straight city road. As we already discussed, this setting occurs sometimes in real life, thus ensuring that the task will not appear artificial to participants. Furthermore, it is still fairly simple, which limits the number of confounding variables that may cause difficulties in interpreting the data.

2.4 Studies Employing Cross-Correlation Function

Cross-correlation is a powerful function which can detect similarities between the given sequences as a function of time or spatial lag applied to one of them. This makes it a versatile tool which has been successfully applied in many fields of science.

An example of its application in marine ecology is the work of Veit et al. [117]. The authors sampled bird abundance and ocean temperature four times a year for eight years off the California shore. They calculated the Pearson correlation coefficients between these two sequences and used a randomization procedure to evaluate statistical significance. The procedure calculated correlation coefficients 100 times between randomly rearranged bird values and original temperatures. For each lag they counted the correlation coefficients from these mismatched sequences that were larger in absolute value than the coefficient calculated using the original matched sequences. If the resulting number was under a threshold, the coefficient calculated using matched sequences was statistically significant. The randomization procedure used by Veit et al. inspired our approach for determining the statistical significance of the obtained cross-correlation results (see Chapter 3 for details).

In neurology, Simpson et al. [118] analyzed the dependence between cerebral blood flow velocity (CBFV) and the power of spontaneous electro-encephalographic (pEEG) signals in healthy term neonates. They calculated the maximum of the cross-correlation function between these sequences for each of their nine participants. In order to test for the statistical significance of the results, they applied a Monte-Carlo method. Namely, using the amplitude spectra of the original signals and randomly generated phase spectra, they calculated uncorrelated CBFV and pEEG signals using an Inverse Discrete Fourier Transform (IDFT). Then they compared the maximum of the true cross-correlation function (obtained using real signals) with the distribution of maxima from the simulated cross-correlation functions. Statistical significance was then determined as the fraction of maxima from the simulated sequences that is larger than the maximum from the original sequences. This way they produced estimates of significance for each subject individually. In contrast, in our approach we provide an overall cross-correlation function estimate as well as its statistical significance level over multiple participants.

Cross-correlation also has its application in time delay estimation (TDE) [119]. TDE is an important research area which has applications in various fields, such as radar, sonar, geophysics, etc. The main goal of TDE is to estimate the time difference that exists between two received signals which are detected by different sensors. If it is the case that the two signals are delayed and attenuated versions of the original signal (such as the echo that can be heard sometimes in the long distance calls), the relative delay between them is equal to the time-lag which maximizes the cross-correlation between these signals. Similarly, in our proposed method we expect that a time lag (delay) exists between glances directed away from the road and increased changes in driving

performance measures (specifically, significant peaks observed in lane position and steering wheel angle cross-correlation functions).

Reich et al. [120] used a cross-correlation function to analyze the spatial relationship between stand characteristics (basal area growth, stand age, site index of productivity, mortality, tree density, number of trees per hectare) of undisturbed, shortleaf pine stands in northern Georgia sampled over two ten-year periods. For each period they calculated the cross-correlation statistic for all pairwise combinations of the above stand characteristics. The results indicated a significant cross-correlation between the basal area growth and other stand characteristics, which were due to small clustering in the northern parts of the state. This was contrary to the regional and broad scale variation that was initially assumed. The authors emphasized the importance of using multiple techniques when interpreting patterns under investigation in order to obtain better understanding. This overall conclusion goes along well with the research presented in this dissertation, since we introduce the cross-correlation measures which can extract important patterns from the driving data in addition to the average-based measures.

Sarvaiya et al. [121] applied normalized cross-correlation function for template matching in medical imaging. Namely, they used small reference images of the areas of interest and detected matching regions in bigger, sensed images. By normalizing the result for the sensed image, they obtained very high recognition rates. The authors concluded that the normalized cross-correlation function provided excellent matching in images both with and without noise. In one case of our method we also propose to normalize the cross-correlation result (hypothesis H2) in order to obtain an estimate of cognitive load changes resulting from individual instances of secondary task engagement.

2.5 Studies Employing Regression Analysis

As stated in the introduction, our third goal (G3) is to provide explanations for the mechanisms underlying our cumulative and instance-based performance measures. We propose in our third hypothesis (H3) that this goal can be accomplished by revealing the variables which significantly contribute to the observed results. To this end, we intend to create multiple regression models which will help in revealing these underlying relationships. This approach has been used often by the researchers in the automotive area and the following paragraphs will review some of their results.

Zhang et al. [30] conducted a driving simulator experiment in order to determine the eye-gaze measures which are diagnostic of decrements in driving performance. The simulated environment comprised of two road types (rural and highway) and two levels of curvature (straight and curvy). As a distraction task, the participants were asked to read common words presented in three rows on displays mounted in the center console, above the dashboard and on the left side of the simulator cabin. The authors derived multiple regression models of the type $Y = a + bX$, where X represented independent variables describing visual attention (such as, total glance duration, weighted gaze variability, weighted gaze vector) and Y represented dependent variables describing driving performance (accelerator release time, standard deviation of lane position and steering entropy). The strengths of the fits, as judged by the coefficient of determination R^2 , ranged from 0.34 to 0.85. Since the slopes of all regression equations were positive, the authors concluded that as the visual distraction increased, driving performance decreased. This agrees with the hypothesis underlying our cross-correlation method that glances directed off-road may negatively impact driving. Similar

to this study, we also intend to use glance duration in our regression analyses. However, we will also include number of glances and PDT away from the road, since they provide additional information about drivers' visual attention.

In the first driving simulator study presented in [15], Horrey et al. explored the impacts of the relative value of tasks (driving and in-vehicle task) and their bandwidths on visual sampling behavior. The value of the task represented which task was prioritized: driving, in-vehicle or both tasks. The bandwidth of the driving task was selected to be low or high by adjusting the frequency of the applied wind gusts. Similarly, the bandwidth of the in-vehicle task was set to either low or high by changing the frequency at which 7-digit phone numbers appeared on an HDD screen. For the in-vehicle task the participants were instructed to read the phone numbers aloud whenever a new number appeared on the screen. A regression equation calculated between the variability of lane position and the mean PDT to the outside world indicated a negative relationship, with PDT explaining 41% of variance encountered in lane position ($R^2 = 0.41$). In other words, as the scanning (PDT) to the outside world decreased, the variability of lane position increased. This conclusion is important and since we intend to include PDT in our regression analyses as well, we expect that our results will point in the same direction: the increase in PDT away from the road should be followed by an increase in our cross-correlation results.

Using the voluntary visual occlusion technique (the participants were instructed to press a button to request a 500msec glimpse of the road) applied in a driving simulator, Tsimhoni and Green [122] examined the visual demand of driving while concurrently interacting with in-vehicle displays. The visual demand of the driving task

was manipulated by driving on roads with four levels of curvature (curve radius). For the secondary task the participants completed a map reading task by responding to questions of varying difficulty. The maps were displayed on an HDD. Regression analysis demonstrated a very strong linear relationship between visual demand and the reciprocal of curve radius ($R^2 = 0.98$). This agreed with the further finding that the mean glance duration towards the in-vehicle screen decreased as the visual demand of driving increased ($R^2 = 0.34$). The overall conclusion was that as the driving visual demand increased, the duration of in-vehicle glances decreased while their number increased. In testing hypothesis H3 we propose to use two “reference” experiments in two different driving environments: highway and city. We expect that the similar result may be obtained in our studies as well.

Another example where regression analysis was successfully applied is a study by Green and George [123] where the authors examined the most appropriate distance from the intersection at which the auditory guidance system should present turn instructions. The experiment was performed in a real vehicle. In one case the participants were following a predefined route and asked when they expected a navigation direction. In the other case, the participants were continuously approaching two different intersections and indicated whether the issued navigation direction was issued too early, too late or about right. Regression analyses revealed a significant effect of the approaching velocity, drivers’ age, direction of turn and gender.

CHAPTER 3

CROSS-CORRELATION METHOD

This chapter provides a detailed description of the cross-correlation method proposed in Chapter 1. The cumulative and instance-based cross-correlation results are demonstrated on two driving simulator studies, which analyze multimodal interactions with two types of in-vehicle devices: PND and MP3 player. The results are compared with the standard average-based measures as well as the subjective measures of cognitive load. Finally, the chapter concludes with the discussion of the observed results.

3.1 Hypotheses Addressed in this Chapter

Our first hypothesis (H1) is concerned with initiator-based quantification of cumulative secondary task engagement. What this means is that it requires an “initiator sequence” which indicates where/when the engagements occur and a “performance sequence” which reflects the effects of those engagements (in our case we are concerned with the effects on driving, although it can be generalized to any other process of interest). Finally, an “extraction function” L is necessary as well which is capable of quantifying the cumulative effect of overall secondary task engagements on driving.

Similarly, our second hypothesis (H2) is concerned with initiator-based quantification of *instances* of secondary task engagement. In this case we intend to estimate the effects of *individual* secondary task engagements on driving and cognitive load. The same aspects discussed in H1 are necessary here as well: extraction function, initiator and performance sequences. However, in this case the extraction function should be adjusted in order to be able to isolate the effects of individual secondary task engagements. The adjustment can be performed by normalizing L with respect to the total number of engagements N : $L' = L/N$.

In both H1 and H2 we proposed to use the mathematical function of cross-correlation as the extraction function L . Cross-correlation function requires two sequences, which agrees with our intention to account for both the initiator and the performance sequence. Detailed explanation of the way cross-correlation function is used in quantifying the cumulative (H1) and instance-based (H2) effects on cognitive load is provided in Section 3.1.3.

Both hypotheses H1 and H2 address the quantification aspects of our first two goals (G1 and G2). However, we also want to be able to rank different types of secondary task engagements based on the results obtained using cumulative and instance-based measures. This is addressed by a common hypothesis H_{RP} and described in Section 3.1.5.

3.1.1 General Terminology

As we indicated in Chapter 1, in this research we are concerned with secondary task engagements which draw visual attention away from the road (visual-only and manual-visual interactions). Therefore, glances directed away from the road are the

obvious choice for the initiator sequence (ρ). In Chapter 2 we defined glances as the general observations of the objects of interest. In that respect they are different from fixations, which are limited in both spatial and temporal domain. However, in this dissertation we are considering *all* glances directed away from the road while the vehicle is moving, which reflects our expectation that they in general negatively affect driving. Nevertheless, since our method is defined in a general fashion, future studies may explore the possibility of using fixations as well.

Regarding a performance sequence (θ), a driving performance measure of interest can be used. In our case we decided to use steering wheel angle and lane position.

The main assumption behind the above choices for the initiator and performance sequences is that any glances directed away from the forward road (as a result of distractions coming from a particular in-car interface) may cause at least a temporary change (worsening is hypothesized) in the driving performance measures. We suspect that this may be the case, because while looking away a driver is not aware of the situation in front of the vehicle, thus making a short pause in willfully controlling a vehicle. Since the situation in front of the car changes dynamically, when visual attention is returned to the road it is likely that the driver will need to perform a correction in order to keep a steady position in the lane. This correction is likely to be correlated with glances returning to the road ahead. The corrections are of course not certain (for example, occasional brief glances at the speedometer may not require corrections). Nevertheless, they are more likely to occur when a driver is occupied with some non-driving related activity (e.g. looking at an HDD). However, if the same trends of driving performance changes keep occurring after looking away from the road, this influence will

be detected by the cross-correlation function. This detection is manifested by the prominent peaks that indicate the position (time lag) where the highest correlation exists between visual attention and a specified driving performance measure, as will be explained shortly. Since our cross-correlation method uses whole sequences, rather than values averaged over long periods of time, it enables us to analyze the experiment in a continuous fashion as time progresses and influences occur.

Let us define two discrete time sequences $\delta[n]$ and $\theta[n]$, which are sampled versions of continuous time signals $\delta(t)$ and $\theta(t)$, respectively. These continuous time signals might represent various processes, but in our case $\delta(t)$ represents gaze angles, while $\theta(t)$ represents a driving performance measure of interest (such as lane position or steering wheel angle). Sampling is performed at some fixed rate, $1/T_s$, where T_s is the sampling period in seconds. Thus, $\delta[n] = \delta(nT_s)$ and $\theta[n] = \theta(nT_s)$.

$\delta(t)$ is sampled by an eye-tracker and is used to obtain a discrete sequence $\delta'[n]$, which contains numerical indexes of the objects that a participant's gaze intersects with. Figure 3.1 shows an example virtual model, which resembles the layout of different objects inside the cabin of our driving simulator. As we can see, various objects are present in the model, such as the speedometer, steering wheel, left and center rear-view mirrors, and so on. The green vector protruding from the yellow avatar indicates the direction of a participant's gaze. Whenever the gaze vector intersects with an object in the virtual model, a corresponding object's numerical index is recorded in the $\delta'[n]$ sequence. In the post-processing we transform $\delta'[n]$ into $\rho[n]$ (initiator sequence), which consists only of 0s and 1s, where 1s indicate glances directed away from the road and 0s indicate glances on the forward road. We consider looking at any of the simulator's

screens (front, left and right screen – blue and green planes in Figure 3.1) as looking at the road, while looking anywhere inside the cabin as away from the road.

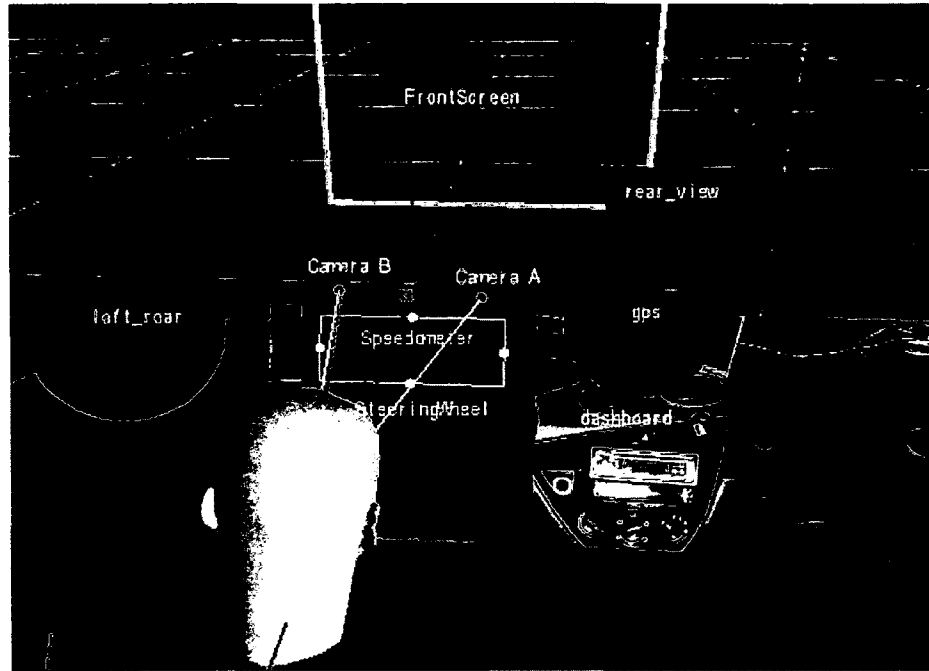


Figure 3.1 Model of our driving simulator's cabin employed in our eye-tracker.

In the following, let us say that $x[n]$ is a sequence of 0s and 1s obtained from the sequence $\rho[n]$, where a 1 represents instants when the driver's gaze returns to the road (after interacting with an in-vehicle device, for example). According to the notation in hypothesis H1, this transformation can be represented as follows: $x[n] = f(\rho[n])$, where f represents a function which extracts the falling edges of the glances, thus producing "reference" points indicating when the gaze returns to the road ahead. Strictly speaking since $\rho[n]$ is a discrete sequence, it is not quite accurate to talk about "falling" edges of the glances, but those are rather the last samples equal to 1 in each glance. Nevertheless, to simplify the terminology, we will refer to the first and the last sample in each glance as the rising and the falling edge, respectively. Figure 3.2 depicts both the

continuous-time and discrete-time representations of x and ρ . We will refer to $x[n]$ as an eye-glance sequence (EGS). Please note that the approach of presenting the data in the continuous-time fashion will be applied to all figures in this dissertation. This significantly improves the visual representation (as can be seen in Figure 3.2); however, we have to keep in mind that all of the variables are in fact discrete sequences.

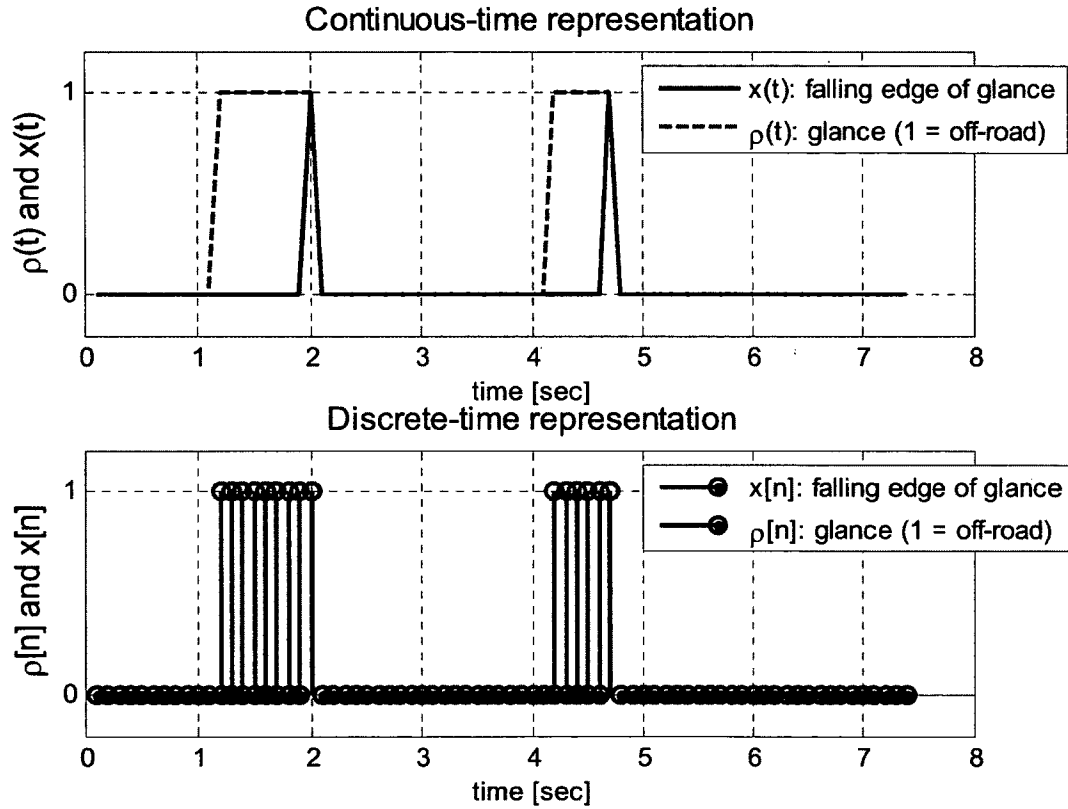


Figure 3.2 Pictorial explanation of the EGS sequence.

Let us also say that $y[n]$ is a measure of driving performance obtained from the raw sequence $\theta[n]$ (e.g., lane position or steering wheel angle). We will refer to $y[n]$ as a driving performance sequence (DPS) and it is obtained by applying some appropriate transformation (g as defined in H1), such as the absolute value of change (AVC), to a driving performance measure of interest ($y[n] = g(\theta[n])$). AVC is defined as follows:

$$AVC\{\theta[n]\} = \frac{|\theta(n) - \theta(n - 1)|}{T_s}$$

Equation 3.1 Absolute value of change (AVC) definition.

and indicates the amount of absolute change in $\theta[n]$ from one sample to another. In this case the larger the value of y , the larger the impact on driving performance.

Based on its definition, AVC always produces positive sequences. In the context of analyzing driving performance, AVC resembles the fact that moving too much towards either side of the road is equally detrimental for driving. AVC provides the magnitude of the change that occurs in a driving performance measure of interest without regard to the direction of the change. It can be argued that the direction of the change is not very important since, if we take city driving as an example, going too far to the right may cause road departure or a collision with parked vehicles, while going too far to the left may cause a collision with the oncoming traffic. The need for corrections (large changes in AVC following the return of visual attention to the road) indicates that something had happened prior to looking back to the road, such as drifting from the center of the lane or an unexpected event (e.g. pedestrian) occurring in front of the vehicle.

3.1.2 Requirements of the Method

Before we continue with the details of the cross-correlation method, it is of interest to discuss three topics that are important for properly preparing the eye-tracker data for the cross-correlation analysis: correcting glance data, filtering glances and sampling rate conversion.

Correcting Glance Data

As we explained in Section 3.1.1 the eye-tracker provides a sequence $\delta'[n]$, which contains numerical indexes (integer numbers ranging from 0 to $K-1$, where K is the total number of objects in the eye-tracker's world model) of the objects that a participant's gaze intersects with. However, it happens occasionally that the eye-tracker does not see the participant's eyes properly and as a result cannot determine where the participant is looking at. This is reported by the index “-1” in the data collection. Some representative examples include when the participant occludes his/her eyes or the eye-tracker cameras with a hand, the head moves too far to the right or to the left thus falling outside of the cameras' field-of-view, and so on. In those situations we have to manually transcribe the data. For this purpose we use videos recorded by the eye-tracker cameras and an additional video of the participant recorded using a separate camera located on the dashboard (see Figure 3.3 below).



Figure 3.3 View of the participant from the camera mounted on the dashboard.

All the videos are recorded simultaneously by off-the-shelf video recording software. Since the eye-tracker overlays a unique frame number on its videos, we can manually go through the data collection and correct the data samples (based on the associated frame numbers) for which eye-tracking was unsuccessful. In rare situations when we are unable to resolve where the participant is looking at, those sections of the data are left unchanged and later are detected and rejected in the cross-correlation analysis. Specifically, experimental segments which contain data samples labeled “-1” are rejected from the further analysis.

Furthermore, it happens rarely (from our experience in less than 2% of experimental segments) that the eye-tracker experiences a temporary delay in collecting the data. This is detected as a “discontinuity” (in other words, a gap) in the time sequence obtained from the eye-tracker. If the discontinuity is long, the information about glance data may be missing. As we will see in section “Sampling Rate Conversion,” we are down-sampling our glance data to 10 Hz, which means that the shortest glance can be 100 msec. This amounts to only 1 sample, so we decided to reject a segment if it contains a discontinuity of at least 2 samples, or 200 msec. Nevertheless, this occurs infrequently.

Filtering Glances

Glances directed away from the road can occur anywhere during the experimental run. Furthermore, due to the very dynamic nature of the eye movements, the eye-tracker measurement errors (according to the manufacturer, a typical error in gaze direction measurement is between 0.5° and 1°) and the gaze instability (which increases as the visibility of eyes decreases), it happens occasionally that the eye-tracker reports glances that are either very short and/or separated by very brief intervals of time (in other

words, just a few samples). The eye-tracker's sampling rate is 60 Hz, which means that the sampling period equals 0.016 seconds. We can ask two questions here: a) how many consecutive samples should be considered to constitute a realistic glance, and b) what is the minimum separation in order for the two glances to be considered as individual glances.

We defined our glances according to Wang et al. [107]: a minimum duration of any individual glance should be 100 msec (which also agrees with SAE J2396 recommended practice [83]) and individual glances should be separated by at least one glance towards a different target (thus, minimum separation is 100 msec). When the gaze travels from the object of interest (e.g., an LCD screen) to another object (e.g., windshield) the eye-tracker noise may appear at the boundary between the two objects. This is detected as a number of very short glances to and from the object of interest. For example, it would appear as if a driver is very rapidly changing the direction of the gaze from the LCD screen to the windshield. Such a rapid change of gaze direction is unrealistic and it can be attributed to the tracking difficulties.

In general, the eye-tracker achieves the best performance when the participant is looking in the general direction of the eye-tracker cameras; however, when the participant changes the direction of the gaze to the side (which is the case when the participants look away from the road towards an LCD screen, dashboard or speedometer), the visibility of the eyes decreases, which contributes to tracking difficulties. Therefore, we can argue that if glances directed off-road appear very close to each other (closer than 100 msec) we can declare that they belong to a single glance. Since those short glances would not be acceptable by the minimum duration rule, we apply the minimum

separation rule first and then the minimum duration rule. This way we can account for those very short glances as well. Nevertheless, if those very short glances are far (>100 msec) from other glances, then we reject those and declare them to be the consequence of the eye-tracking imperfections. Figure 3.4 illustrates the above procedure.

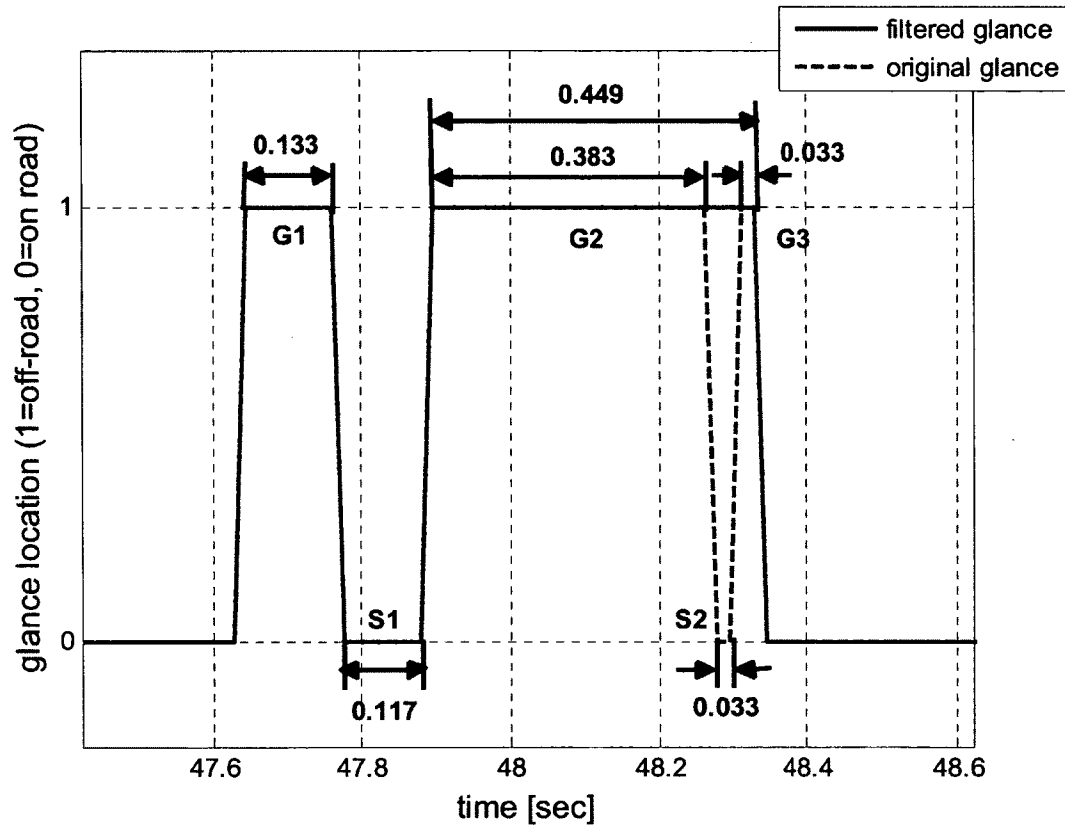


Figure 3.4 Illustration of the glance filtering procedure.

The red dotted line in Figure 3.4 represents the original glance sequence reported by the eye-tracker, while the solid blue line represents the filtered glance. We can see that there are 3 glances away from the road in total reported by the eye-tracker: G1, G2 and G3. Their durations are 0.133, 0.383 and 0.033 seconds, respectively. According to the minimum duration rule (≥ 100 msec), we would have to reject G3. However, the separation between G2 and G3 ($S2 = 0.033$ msec) is less than 100 msec.

Thus, we concatenate G2 and G3 into a single glance. Since both the separation S1 (0.117 seconds) between G1 and G2 and the duration of S1 (0.133 seconds) are longer than 100 msec, we can accept G1 as being an individual glance. As the final result, we obtain a filtered glance sequence which consists of only two glances (solid blue line).

Sampling Rate Conversion

In general, separate equipment is used for obtaining driving performance and visual attention data. Therefore, different sampling rates may be employed. Specifically, in the case of driving data (such as steering wheel angle, lane position, throttle position, and velocity) typical sampling rates found in the literature range from 5 Hz to 50 Hz [13;29;50;53;107;122;124;125], while in the case of visual attention data (such as gaze angles, pupil diameter, and blinking) sampling rates range from 30 Hz to 60 Hz [29;50;107;124]. Even though our driving simulator supports higher sampling rates, we collected all driving related data at 10 Hz for two reasons. First, 10 Hz is commonly used in the literature [107;124;125]. And second, very high sampling rates have very short sampling periods during which not enough change accumulates between consecutive samples to be detected by our driving simulator. Namely, the steering wheel angle is changing relatively slowly (the majority of our experiments are conducted on straight roads or straight sections of city roads) and the resolution of the rotational encoder used for obtaining the steering wheel angle is limited to 0.1° . As a result, when the AVC transformation is applied to steering wheel angle, the observed changes between consecutive samples equal either 0° or 0.1° . This way we obtain a binary variable, which is not useful for determining where the largest changes occur. On the other hand, a 0.1 second interval allows enough change to accumulate.

The eye-tracker data was collected at 60 Hz, which is the only available rate offered by the eye-tracker (see Appendix B for a detailed overview of the eye-tracker's capabilities). In order for both EGS and DPS sequences to represent the system in the same fashion, they must be sampled at the same rate. This is accomplished by down-sampling the eye-tracker data from 60 Hz to 10 Hz. However, due to differences in time when the initial sample was taken, jitter in sampling and so on, the samples from both sequences do not have to occur at the same time instants. In other words, each device has its own time scale. The synchronization of the zero points of the two time scales is performed by issuing synchronization signals by a custom software/equipment at the beginning of each experiment. Those synchronization signals are then detected on both devices and used as zero points. Appendix A provides a detailed overview of the synchronization procedure. Even though the zero points are synchronized, we cannot perform a simple down-sampling by just keeping every sixth sample from the original eye-tracker data ($60 \text{ Hz} / 10 \text{ Hz} = 6$). Instead, we apply the following custom procedure:

- a) Detect time instants when each glance starts and ends in the 60 Hz time scale (“rising” and “falling” edges) in the glance location sequence ($\rho[n']$) obtained from the eye-tracker.
- b) For each edge in the 60 Hz time scale, find the closest time instant in the simulator's 10 Hz time scale and make it the new edge.
- c) Initialize all samples (now in the 10 Hz time scale) between the new edges to 1s, and all remaining samples to 0s.

The above procedure produces a sequence of glances aligned to the driving simulator's time scale (10 Hz). Figure 3.5 shows one specific example based on actual

data. Please note that the signals depicted in this figure are discrete. However, in the interest of better visual representation, we plotted both signals as continuous functions, rather than individual samples. 1s indicate glances directed off road, while 0s indicate glances directed to the road ahead. The solid blue and dashed red lines show locations of glances represented on the 60 Hz and 10 Hz time scales, respectively.

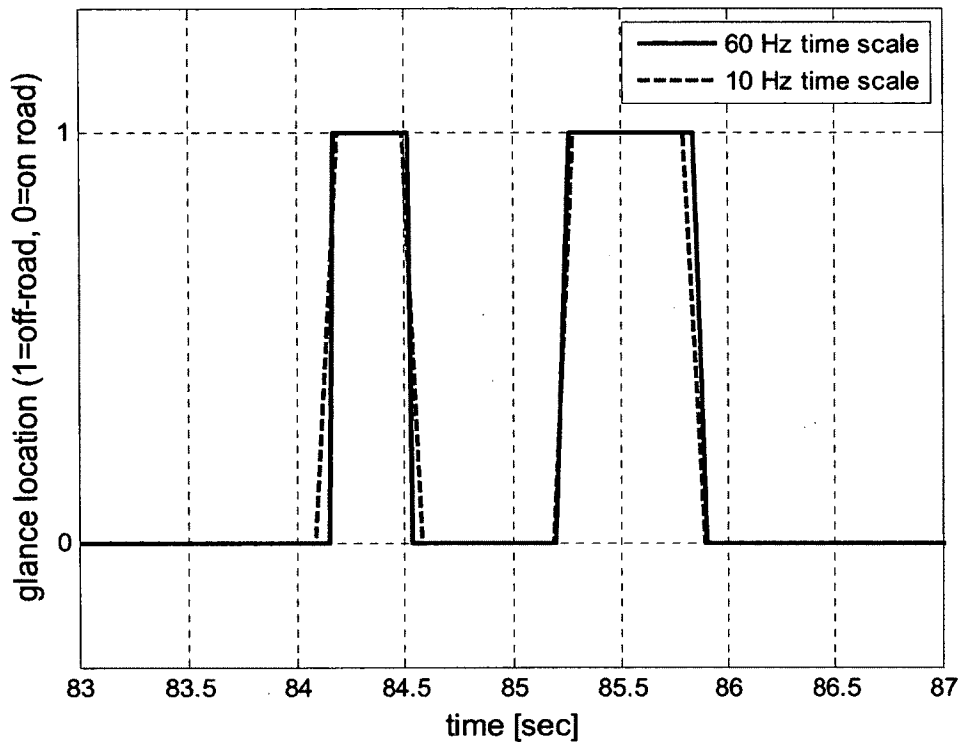


Figure 3.5 Converting glances from the 60 Hz time scale (eye-tracker) to the 10 Hz time scale (driving simulator).

We can see in Figure 3.5 that there exists a slight mismatch between the glances presented on two time scales. This is expected, since the samples on two scales do not have to be aligned. Figure 3.6 shows the zoomed-in falling-edge of the first glance from Figure 3.5. As we can see, the edge of the glance on the 60 Hz time scale falls at 84.52 seconds, which is between 84.49 seconds and 84.59 seconds on the 10 Hz scale. If

we check the differences we can see that the smallest one of 0.03 seconds is obtained if we take 84.49 seconds to be the falling edge of the glance on the 10 Hz time scale.

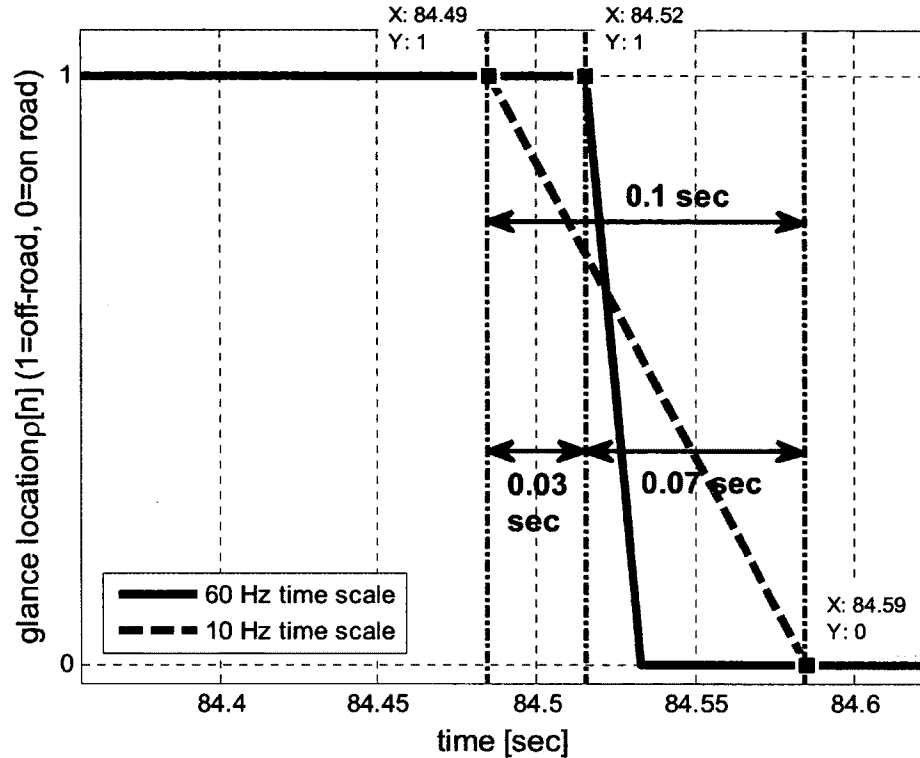


Figure 3.6 Searching for the closest sample on the 10 Hz time scale.

By finding the closest time when converting glances from one scale to another, we obtain the best conversion, as opposed to only taking the times larger or smaller than the reference time on the 60 Hz scale (“rounding” up or down). This way the maximum theoretical conversion error equals to ± 0.05 seconds, which occurs when the edge of the glance in the 60 Hz time scale falls exactly between two consecutive samples in the 10 Hz time scale. However, we wanted to empirically check the error which is introduced in the process. For this purpose we calculated time differences in rising edges of glances in 60 Hz and 10 Hz scales for one of our studies that will be presented in Section 3.2.2. We repeated the same procedure for the falling edges of glances as well. Figure 3.7 shows the

histograms of time differences obtained in each case. In creating these histograms we used data from 12 participants, which amounted to a total of 2536 glances. We can see that for both rising and falling edges there is practically a uniform distribution of time differences around the actual time obtained by the eye-tracker (0 seconds mark in both histograms). Furthermore, we conducted a non-parametric Kolmogorov-Smirnov test in order to confirm the above qualitative explanation. The test revealed no significant deviations from the uniform distribution in each case: $p=0.23$ for falling edges and $p=0.95$ for rising edges. These results indicate that there is no bias towards any direction (left or right from the reference) when performing the conversion. Also, we can see that in both cases a maximum offset from the actual glance edge is equal to ± 0.05 seconds, which confirms the expected maximum conversion error.

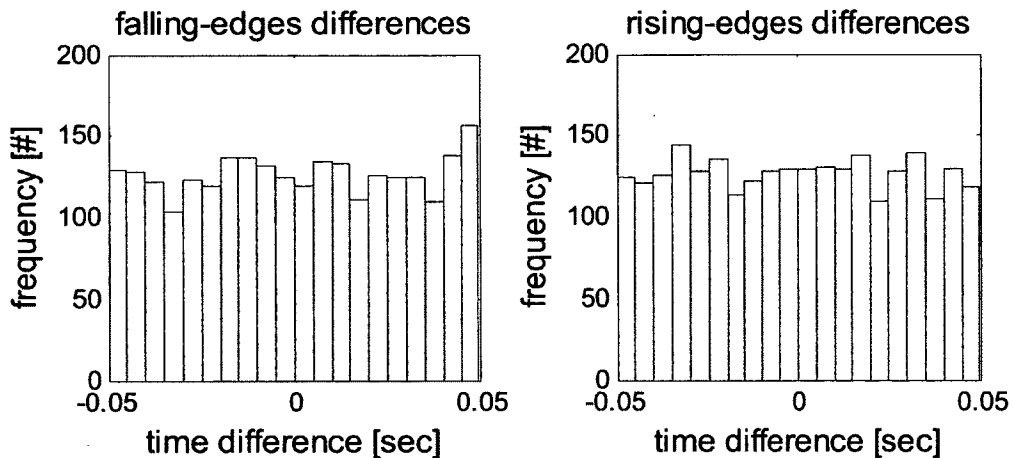


Figure 3.7 Differences in rising and falling edges between two time scales.

3.1.3 Definition of the Method

Cross-correlation function can be used to indicate an association between two sequences. The association may emerge due to a relationship between the sequences that

may be either causal or indirect through some known and unknown mechanisms. In the case of driving, the relationship between an EGS sequence $x[n]$ and a DPS sequence $y[n]$ may exist due to the need for correcting the car's position in the lane after returning the gaze to the road. The following sections provide definitions for the cumulative and instance-based quantifications of secondary task engagements based on cross-correlation.

Initiator-based Quantification of Cumulative Secondary Task Engagement

As proposed in hypothesis H1, the initiator-based quantification of cumulative secondary task engagement can be performed by cross-correlating EGS and DPS sequences. For two discrete time, causal sequences $x[n]$ and $y[n]$ of equal and finite length N , with (possibly) non-zero values for $0 \leq n \leq N - 1$, cumulative effect of secondary task engagement ($R_{xy}[lag]$) can be estimated as follows [126]:

$$R_{xy}[lag] = \begin{cases} \sum_{m=0}^{N-1-lag} x[m]y[m+lag], & 0 \leq lag \leq N-1 \\ \sum_{m=-lag}^{N-1} x[m]y[m+lag], & -N+1 \leq lag < 0 \end{cases}$$

Equation 3.2 Initiator-based quantification of cumulative secondary task engagement using a cross-correlation function between discrete sequences $x[n]$ and $y[n]$.

If $x[n]$ is a non-negative sequence of 0s and 1s, and $y[n]$ is a non-negative measure of driving performance (such as the AVC of lane position), $R_{xy}[lag]$ will always be greater or equal to zero. In such a case, a peak in $R_{xy}[lag]$ might indicate that changes in $x[n]$ (initiator sequence) are associated with changes in $y[n]$ (performance sequence) that occur after a certain number of samples, lag . In fact, the indication is that changes in $\rho(t)$ are associated with changes in $\theta(t)$ after a time period ΔT , where ΔT is

related to the sampling period T_s and the *lag* as $\Delta T = \text{lag} \cdot T_s$. The larger the peak in $R_{xy}[\text{lag}]$, the higher the association. Also note that, for causal sequences $x[n]$ and $y[n]$ that are of equal and finite length N , R_{xy} will have at most $(2N - 1)$ values. The cross-correlation function obtained using the Equation 3.2 tells us how large the *cumulative (overall)* effect is on the change of a driving variable of interest when looking away from the road ahead. Specifically, each time a glance directed off-road appears, there is a 1 in the $x[n]$ sequence. This results in products which are added to the total sum. In this respect it is similar to variance, since it characterizes driving performance in each segment as a whole.

In general, the experimental segments can be of different length. In such a case, we may want to introduce weighing for the cumulative cross-correlation functions. The reason is that the glances which occur over a short segment should have higher importance than the glances occurring over a long segment, which agrees with our argument from the introduction that more frequent interactions may produce larger effects on driving. The weighing can be accomplished based on each segment's length relative to the total length of all segments. If this weighing is desired, the following equation should be used instead of Equation 3.2:

$$R_{xy}^k[\text{lag}] = \begin{cases} \frac{T_k}{T} \sum_{m=0}^{T_k-1-\text{lag}} x[m]y[m+\text{lag}], & 0 \leq \text{lag} \leq T_k - 1 \\ \frac{T_k}{T} \sum_{m=-\text{lag}}^{T_k-1} x[m]y[m+\text{lag}], & -T_k + 1 \leq \text{lag} < 0 \end{cases}, \text{ with } T = \sum_{k=1}^M T_k$$

Equation 3.3 Initiator-based quantification of cumulative secondary task engagement with weighing which accounts for the segment length.

where T_k is the length of the k^{th} segment ($k = 1, \dots, M$), M is the total number of segments and T is the total length of all segments taken together. In the studies presented at the end of this chapter the weighing was not necessary, since all segments were of the same length.

Initiator-based Quantification of Instances of Secondary Task Engagement

Hypothesis H2 proposes to estimate the amount of change *per individual instance of secondary task engagement* and is defined according to Equation 3.4. N_g is the total number of instances of secondary task engagement (in our case, glances directed off-road on a corresponding segment). This way we are able to estimate on average how detrimental each individual glance is to driving.

$$R_{xy}[lag] = \begin{cases} \frac{1}{N_g} \sum_{m=0}^{N-1-lag} x[m]y[m+lag], & 0 \leq lag \leq N-1 \\ \frac{1}{N_g} \sum_{m=-lag}^{N-1} x[m]y[m+lag], & -N+1 \leq lag < 0 \end{cases}, \text{ with } N_g = \sum_{n=0}^{N-1} x[n].$$

Equation 3.4 Initiator-based quantification of instances of secondary task engagement using a normalized cross-correlation function between discrete sequences $x[n]$ and $y[n]$.

Naming Conventions

We will make a couple of naming conventions here which will simplify the terminology in the rest of the text. From now on we will refer to the “cumulative secondary task engagement cross-correlation results” as the *cumulative cross-correlations*. Similarly, “instance-based secondary task engagement cross-correlation results” will be referred to as the *per-glance cross-correlations*.

As we explained in Section 3.1.1 our driving performance sequence ($y[n]$) is obtained by applying the AVC transformation to the driving performance measures of interest. In our case, we decided to use steering wheel angle and lane position. Of course, a separate DPS sequence is obtained for each measure. Those DPS sequences are then cross-correlated with the EGS sequence to obtain cumulative and per-glance results. In order to simplify the terminology, we will refer to the “cumulative cross-correlation results obtained using the absolute value of change of steering wheel angle” as the “cumulative steering wheel angle cross-correlations.” The same convention will be used for lane position. Likewise, “per-glance cross-correlation results obtained using the absolute value of change of steering wheel angle” will be referred to as the “per-glance steering wheel angle cross-correlations.” An abbreviation that will be used in the graphs is R_* , where “*” refers to either steering wheel angle (*sw*) or lane position (*lp*). We do not introduce any additional sub- or superscripts in this abbreviation to distinguish between cumulative and per-glance cross-correlation results, since this distinction will always be clear from the context.

Example Cumulative and Per-Glance Cross-Correlation Results

In order to make the cross-correlation calculations easier to comprehend, we will present a simple artificial example depicted in Figure 3.8.

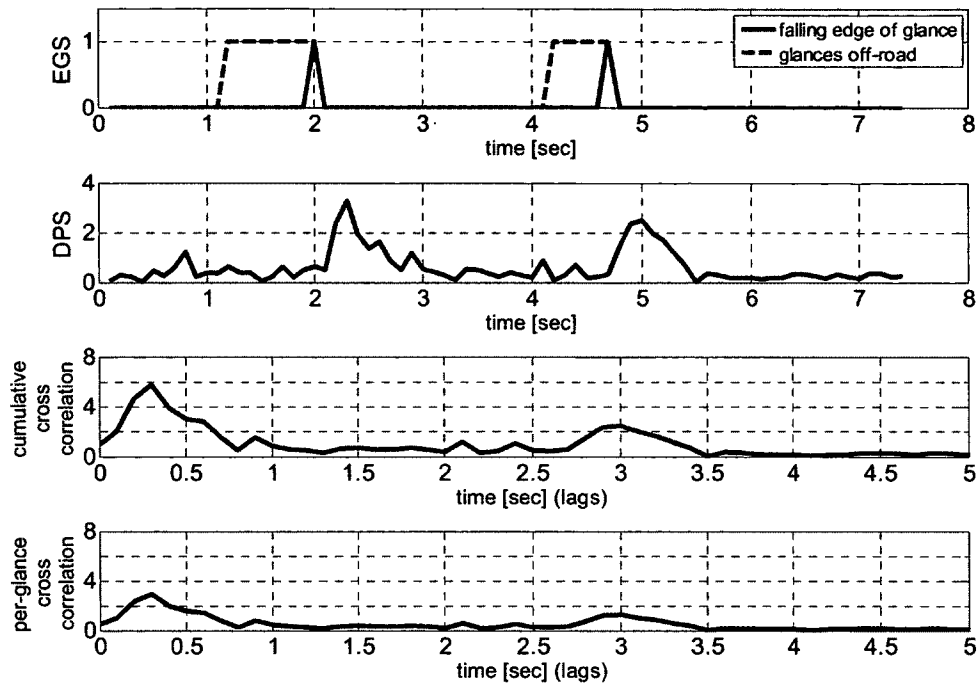


Figure 3.8 Example cumulative and per-glance cross-correlation results using two simulated sequences.

As before, all variables presented here are discrete, however, continuous representation makes them easier to comprehend visually. The figure shows four graphs. The top graph indicates that the driver looks away from the road twice (dashed red line). The solid blue line in the top graph is the EGS sequence (initiator sequence) and it consists of 0s everywhere, except where the driver's gaze returns to the road. The second graph shows a DPS sequence which was obtained using the AVC transformation. For the purpose of this example, we selected the two major changes in DPS to appear 0.3 seconds after the impulses in the EGS sequence. This indicated the hypothesized corrective actions following the two glances directed away from road. Finally, the two bottom graphs show the cumulative and per-glance cross-correlation results obtained for the above sequences.

Since in this case we are using only a single experimental segment, it is very easy to relate the cumulative and per-glance cross-correlation results. Namely, the per-glance result is essentially a normalized version of the cumulative result, with the normalization factor being equal to $1/N_g$, where $N_g = 2$ is the total number of glances in this segment (compare Equation 3.2 and Equation 3.4). As we will see in the next section, in actual studies we calculate both cumulative and per-glance cross-correlation results for many experimental segments and participants, which are then averaged to obtain one overall response.

As discussed before, the EGS sequence consists of 1s where the gaze returns to the road ahead and 0s everywhere else. This is indeed an initiator sequence, since those 1s are used as reference points in the cross-correlation formula. Hence, we can say that this sequence represents a dimensionless quantity. On the other hand, DPS sequence is in our case obtained by applying the AVC transformation to either steering wheel angle or lane position, which makes its units [degrees/second] or [meters/second], respectively. Since the cross-correlation formula involves multiplication of the samples from the EGS and DPS sequences, we can say that the units for the cumulative cross-correlation result are [degrees/second] for steering wheel angle and [meters/second] for lane position. The same units essentially apply in case of the per-glance cross-correlation result. However, it is important to emphasize that this result is based on the instance of secondary task engagement, that is, per-glance.

As we can see in Figure 3.8, the highest cross-correlation peaks appear at the lag of 0.3 seconds (after the gaze returns to the road) for both cumulative and per-glance cross-correlation functions. This lag is equal to the separation between falling edges of

glances (in EGS sequence) and the observed peaks in changes in driving performance (in DPS sequence). The second largest peak appears at the lag of 3 seconds and it is the result of the first glance (occurring at 2 seconds) getting correlated with the second change in DPS (occurring at 5 seconds). This is the result of the way cross-correlation function is calculated. Namely, one sequence is being shifted over the other one, thus the changes in driving performance may get correlated with glances which are not directly related to them. These “distant” correlations occur far from the lag of zero and are smaller in magnitude, since a smaller number of glances contributes to those. In contrast, both glances contribute to the largest peak, since each of those was followed by a large change in the DPS sequence. This way, if the changes in driving performance typically occur at similar distances following the glances away from the road, this effect will be detected by the cross-correlation function. It is also worth noting that correlations of “distant” glances and changes in driving performance do not pose problems. The reason is that in reality the glances do not occur at the same locations, thus the influences of any “distant” correlations will be dispersed over many different lags and eventually eliminated (or at least attenuated) when multiple cross-correlation functions are averaged over multiple segments (as will be presented in the next section).

3.1.4 Algorithm

When performing driving-related experiments (although it can be generalized to any other type of experiment) the following requirements are needed for any estimation procedure:

1. Should be performed over multiple participants.

2. Should be performed over multiple segments of road (or experiment epochs).
3. Should provide estimates of statistical significance.

The rest of this section addresses these requirements and presents the algorithm for implementing the proposed cross-correlation method.

Estimating Cross-Correlation Results

First, note that in the following, whenever possible, we will drop the discrete time variables lag and n . For example, $R_{xy}[lag]$ and $x[n]$ will become R_{xy} and x , respectively.

In order to estimate R_{xy} using either Equation 3.2 (cumulative) or Equation 3.4 (per-glance), let us consider sequences x_{ij} and y_{ik} . The subscript i designates the participant ($i = 1, \dots, K$) who generated the data. The subscript j designates the segment on which the glance data was collected ($j = 1, \dots, M$). The subscript k designates the segment on which the driving performance data was collected ($k = 1, \dots, M$). When calculating R_{xy} both x_{ij} and y_{ik} sequences have to originate from matched segments, thus $j = k$. For a particular participant and segment, a peak in R_{xy} at time $\Delta T = lag \cdot T_s$ can indicate deterioration in driving performance following the return of the gaze to the road.

Once we calculate R_{xy} for each participant and each segment we can turn to requirements 1 and 2 outlined above. To meet these requirements, we average the results of the cross-correlation calculation over all participants (requirement 1) and all segments (requirement 2) and of course we do this for each value of lag . Note that averaging has to

take into account that segments may potentially be of different length N . We only average R_{xy} for values of *lag* that can be estimated for all segments.

Figure 3.9 shows a pseudo-code (P-C.1) that implements the algorithm described above. First, we introduce the segment pointer sequences S_x and S_y , which are used to select x and y from the appropriate road segments. For matched segment calculation, the two pointer sequences are the same and they select consecutive segments. Next, we calculate the cross-correlation functions for each participant and each segment using either Equation 3.2 or Equation 3.4 and then average the results over all participants and segments. Hence, we obtain one global cross-correlation function for all lag values of interest. When calculating the cumulative cross-correlation function, if the lengths of segments are different, Equation 3.3 can be used instead of Equation 3.2, since it introduces appropriate weighing for the segment length.

```

P-C.1: Pseudo-code for calculating cross-correlation results  $R_{xy}$ 
// cross-correlation averaged over all segments and participants

segment pointer sequence  $S_x = \{1, \dots, M\}$ 

if matched segments for  $x$  and  $y$ 
    segment pointer sequence  $S_y = S_x$ 
else if mismatched segments for  $x$  and  $y$  (for statistical significance calculation,
                                         see pseudo-code in Figure 3.10)
    segment pointer sequence for  $S_y = \text{permute}(S_x)$ 
end

for each participant  $p_i, i = 1, \dots, K$ 
    for each segment  $j = 1, \dots, M$  pointer from sequences  $S_x$  and  $S_y$ 
         $R_{ij} = \text{xcorr}(x_{iS_x(j)}, y_{iS_y(j)})$  //apply either Equation 3.2, Equation 3.3 or Equation 3.4
    end
end

 $R_{xy} = \text{average over all } i, j \text{ of } R_{ij}$ 

```

Figure 3.9 Pseudo-code for estimating cross-correlation results.

Estimating Significance of Cross-Correlation Results

Our cross-correlation tool would not be very useful if it did not also estimate the statistical significance of its output (requirement 3). To this end we will use a randomization procedure similar to that employed by Veit et al. [117] as described in Section 2.4. The same procedure applies to both cumulative and per-glance cross-correlation results.

In testing statistical significance, our null hypothesis is that the values of the cross-correlation function at a particular lag ($R_{xy}[lag]$) calculated using matched segments are due to chance. We can test this hypothesis by comparing the values $R_{xy}[lag]$ to many (e.g. P) cross-correlations between sequences with characteristics similar to x and y , but without any association to each other. To this end in the randomization process we use x and y sequences from mismatched segments. This approach produces sequences with identical characteristics to the ones used to calculate $R_{xy}[lag]$. Also, barring a problem with our experimental design, the P calculations of R_{xy} on mismatched segments are the results of chance and should indicate no association between the sequences. Realizing that larger cross-correlation magnitudes indicate higher association between the sequences, we can estimate the statistical significance of $R_{xy}[lag]$ based on how its magnitude compares to the magnitudes of the P values calculated using mismatched segments.

Thus, our randomization procedure compares cross-correlations between eye-glance and driving performance sequences (x_{ij} and y_{ik}) on matched segments ($j = k$) to those on mismatched segments ($j \neq k$). We can calculate R_{xy} many (e.g. $P = 1,000$) times using mismatched segments. Let us designate the resulting P sequences as R_m ,

$m = 1, \dots, P$. We are interested in comparing the magnitude of $R_{xy}[lag]$ (which was calculated using matched segments) to the P $R_m[lag]$ values (mismatched segments). If $R_{xy}[lag]$ is larger than the $(\alpha \cdot P)^{th}$ -largest $R_m[lag]$ (we will refer to this as $R_{sig}[lag]$), we can claim that $R_{xy}[lag]$ is statistically significant with $p < \alpha$, where α is a desired significance level and $0 < \alpha < 1$. Thus, in rejecting the null hypothesis (which proposes that our estimate of $R_{xy}[lag]$ is due to chance), the probability of making a Type I error is less than α .

As an example, for $P = 1,000$ and $\alpha = 0.05$, if $R_{xy}[lag]$ is larger than the 50th $R_m[lag]$ value, we can claim that $R_{xy}[lag]$ is statistically significant with $p < 0.05$ (for this example, $\alpha \cdot P = 0.05 \cdot 1,000 = 50$). This is because our calculations of mismatched cross-correlations, which represent outcomes based on chance, produced magnitudes that are larger than or equal in magnitude to our $R_{xy}[lag]$ in less than 5% of the cases (at most 49 out of 1,000).

Figure 3.10 introduces pseudo-code (P-C.2) for calculating the values of R_{sig} that designate the margin above which a value for R_{xy} can be considered statistically significant. For each value of lag the code arranges P cross-correlation values calculated using mismatched segments into descending order. This produces the sequences $O_o[lag]$, $o = 1, \dots, P$, with $O_1[lag]$ being the largest. Using $O_o[lag]$ we can easily find the $(\alpha \cdot P)^{th}$ -largest value for $R_m[lag]$: it is $O_q[lag]$ where $q = \alpha \cdot P$. Note that P-C.1 uses the segment pointer sequences S_x and S_y to create mismatched sequences. The sequence S_y is a permuted version of the sequence S_x , with a different permutation for each $m = 1, \dots, P$.

For example, for $P = 1,000$ and a desired significance of $\alpha = 0.05$ we need to set $R_sig[lag]$ to the value of the 50th largest $R_m[lag]$ (because $0.05 \cdot 1,000 = 50$). Thus, $R_sig[lag] = O_{50}[lag]$. If $R_{xy}[lag]$ is larger than $O_{50}[lag]$, $R_{xy}[lag]$ is statistically significant with $p \leq 0.05$.

```

P-C.2: Pseudo-code for calculating statistical significance
// values of  $R_{xy}(lag) > R\_sig(lag)$  are statistically significant

 $P$  = number of mismatched cross-correlations
set  $\alpha$ ,  $0 < \alpha < 1$ 
set index  $q = \alpha * P$ 

calculate  $P$  mismatched cross-correlations  $R_m$ ,  $m = 1, \dots, P$ , using P-C.1 (Figure 3.9)

for each  $lag$ 
     $O_o(lag)$  = order values of  $R_m(lag)$  in descending order of magnitude,  $o=1, \dots, P$ ,  $m = 1, \dots, P$ 
     $R\_sig(lag) = O_q(lag)$  //  $q$  is index based on significance level  $\alpha$ 
end

```

Figure 3.10 Pseudo-code for calculating statistical significance.

3.1.5 Ranking Cross-Correlation Results

The calculations of cumulative and per-glance cross-correlation results presented in the previous section can be applied to studies with either a single or multiple experimental conditions. In case of a single experimental condition, a significant cross-correlation peak indicates the presence of the effect of looking away from the road on driving. Similarly, the same effect can be analyzed for multiple experimental conditions, where each would have a corresponding cross-correlation function and an estimate of significance. However, as proposed in hypothesis H_{RP} , it may also be of interest to analyze whether the experimental conditions are significantly different from each other with respect to their cumulative and per-glance results, which would allow ranking.

According to H_{RP} , this section presents two approaches that we can be taken here. We will explain each approach individually and then present a pseudo-code which will demonstrate how they can be used (see Figure 3.11). It can be argued that either approach is equally valid, thus observing the conclusions from both is likely the best solution.

Extracting the Magnitudes of the Most Prominent Peaks

The first approach is concerned with the difference that may exist specifically between the most prominent cross-correlation peaks (see for example Figure 3.20, pg. 136). To generalize the approach presented in P-C.1 (Figure 3.9), besides the number of participants K and the number of road segments M , let us assume that there are also L experimental conditions. Then, we can symbolically present the collection of all cross-correlation functions (for all participants, segments and conditions) as R_{ij}^l , where $i = 1, \dots, K$, $j = 1, \dots, M$ and $l = 1, \dots, L$. Equivalently, we can present the overall, average cross-correlation function for each experimental condition as R_{xy}^l . According to the algorithm presented in the previous section, each cross-correlation function for each experimental condition (R_{xy}^l) is obtained by averaging (per lag) a family of curves calculated for individual participants and road segments. This means that for any lag we can isolate a separate group of samples that belongs to each experimental condition. Specifically, we find a *lag* which corresponds to the most prominent peak in each of L final cross-correlation functions (R_{xy}^l) and isolate up to $K \cdot M$ samples (number of participants times the number of road segments) for each experimental condition ($R_{ij}^l[lag]$). This gives us L groups of samples which can be compared statistically.

Extracting the Areas below the Cross-Correlation Curves

The second approach is concerned with the difference that exists between experimental conditions over a range of lags as opposed to looking at a single lag. It can be termed as the “area below the curve” approach and was inspired by Strayer and Drews [60]. Namely, the authors quantified the amplitude of the P300 component of event-related brain potentials for two experimental conditions by calculating the area below each P300 function. The area was calculated for a time interval which included the largest change in P300. This procedure was performed for each participant/experimental condition. Finally, they statistically compared the calculated areas between the two experimental conditions. Similar approach can be applied in our case as well. Namely, for a desired range of lags $[lag_{start}; lag_{end}]$ we can calculate the area below the cross-correlation functions for each combination of experimental conditions, participants and segments ($R_{ij}^l[lag_{start}; lag_{end}]$). As with the previous approach, this provides us with L groups of areas below the curves which can be compared statistically.

Common Statistical Analysis

Previous two subsections presented two approaches which enable characterizing the cross-correlation results for each experimental condition. In this section we perform statistical analyses in order to evaluate whether significant differences exist between the experimental conditions.

A data collection which holds either the magnitudes of the most prominent cross-correlation peaks or the areas below the curves for the specified range of lags can be symbolically presented as C^l . For each $l = 1, \dots, L$ this data collection should contain up to $K \cdot M$ entries. Once we have the data divided into separate groups (conditions), we

can perform statistical comparisons between them, as proposed in H_{RP} . In order to make this approach as universal as possible, we decided to use the Kruskal-Wallis test, which is a non-parametric version of the classical one-way ANOVA and an extension of the Mann-Whitney U test to more than two groups (since, in general, L can be larger than 2). This way the procedure does not depend on the assumptions underlying the parametric methods and can accept the data which is not normally distributed, which makes the procedure applicable to a larger number of cases. If the Kruskal-Wallis test shows the existence of the main effect, we also perform post-hoc pairwise comparisons using the Wilcoxon test in order to determine the experimental conditions that are different from each other. If the pairwise comparisons demonstrate significant differences as well, we can conclude that the observed cross-correlation results are not only significant individually, but that they are also significantly different from each other. This ranking procedure can be applied to both cumulative and per-glance cross-correlation results.

Figure 3.11 shows the pseudo-code which algorithmically outlines the steps described in the previous paragraphs.

P-C.3: Pseudo-code for testing statistical difference between experimental conditions

```
//set these two variables if the desired approach is area under the curves
set  $lag_{start}$ 
set  $lag_{end}$ 

for  $l = 1, \dots, L$ 
     $lag = \text{find the lag of the most prominent peak for } R_{xy}^l$ 
    for  $i = 1, \dots, K$ 
        for  $j = 1, \dots, M$ 
            if Approach == 'compare peaks'
                set  $peak = R_{ij}^l[lag]$ 
                append  $peak$  to  $C^l$ 
            else if Approach == 'compare areas'
                set  $area = \text{calculate area below } R_{ij}^l[lag_{start}:lag_{end}]$ 
                append  $area$  to  $C^l$ 
            end
        end
    end
end

main_effect = Kruskal-Wallis ( $C^l$ ),  $l=1, \dots, L$ 
post_hoc = Wilcoxon pairwise comparisons ( $C^l$ ),  $l=1, \dots, L$ 
```

Figure 3.11 Pseudo-code for testing statistically significant differences between experimental conditions.

3.2 Studies Implementing Cross-Correlation Method

This section gives a detailed description of two driving simulator studies which were used for testing hypotheses H1, H2 and H_{RP} proposed in the introduction. For each study we present both the results obtained using our cross-correlation method, as well as using the average-based measures. This allowed for direct comparison between the two approaches. Furthermore, we analyze the subjective estimates of cognitive load and provide comparisons with those as well.

3.2.1 Exploring Augmented Reality Navigation Aids

This study ([36] © 2011 Association for Computing Machinery, Inc. Reprinted by permission) was oriented towards predominantly visual in-car interactions (listening to voice directions was also involved) and it compared a standard map-based PND (SPND) with two emerging navigation aids: augmented reality (AR) and street view (SV).

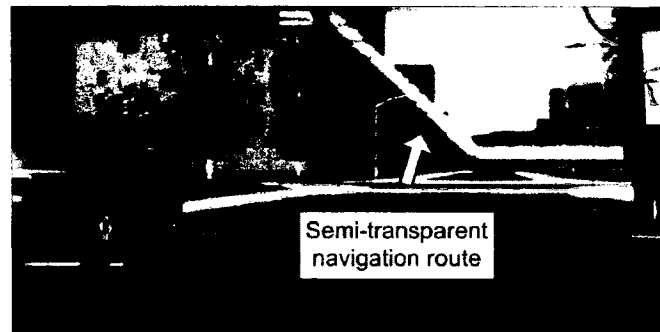


Figure 3.12 AR navigation aid shown from driver's perspective.

AR (Figure 3.12) overlays a semi-transparent navigation route directly on the windshield, thus not requiring drivers to take their eyes off the road in order to obtain navigation information. In our driving simulator, the navigation route was projected onto simulator screens, which created an illusion of it being displayed on the windshield. Therefore we can say that it uses full windshield as a head-up display (HUD). The navigation route was suspended above the center of the road at a height of about 2 meters. This produced the visual effect of a navigation route hovering above the vehicle, similar to the Virtual CableTM [127].

SV (Figure 3.13) navigation has been made possible through the proliferation of smart phones and online resources, which provide street level views of the roads (similar to the Google Street View [128]).



Figure 3.13 SV navigation aid displayed on LCD.

It presents a sequence of images of the surrounding world taken from the driver's perspective (egocentric view). This sequence is augmented with a translucent, wide, road-level surface which represents the navigation route. We decided to use this road-level surface because of its similarity to commercially available HDD-based PNDs (such as [128]). The images were shown on a head-down display (HDD) and they were changing as the driver advanced through the world. In reality, SV would use images taken at a prior time. This was faithfully simulated in our study by another driving simulator which was running in parallel with the one operated by the participants. Specifically, static entities (such as signs and buildings) were the same in both simulations, while the vehicles (parked and moving) and pedestrians were different. A new image was displayed on LCD every 15 meters, which is approximately the distance used in Google Street View.

Finally, SPND (Figure 3.14) represents a common map-based navigation device with an exocentric, "top-down" view. It was also presented on an HDD. The small green triangle visible in Figure 3.14 indicates the current position of the car and it always remained in the center of the screen, while the map rotated about it. The pink line indicates the navigation route. The main reason for including an SPND in our experiment

is their common presence in vehicles nowadays. Thus, any observed differences between the PNDs on test would be the easiest to characterize with respect to SPND.



Figure 3.14 SPND navigation aid displayed on LCD.

Since most contemporary PNDs enable voice directions, we decided to include identical turn-by-turn directions for all three PNDs in our experiment. The directions were prerecorded by a voice talent in order to eliminate potential problems with the comprehension of synthesized speech [14].

Figure 3.15 shows how the experimental setup looked like inside the cabin.

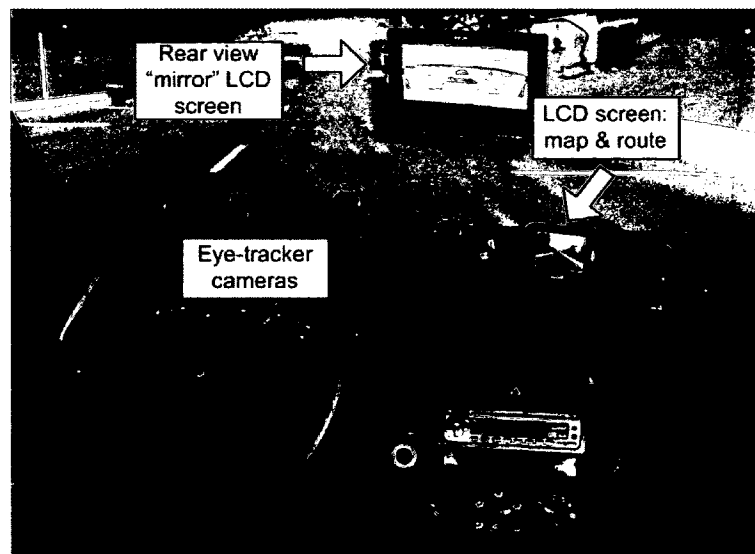


Figure 3.15 Experimental setup inside the simulator cabin.

The LCD screen (HDD) was placed on top of the dashboard and it was used by both SV and SPND navigation aids. The eye-tracker was also used in this study. A camcorder was installed on the far right side of the dashboard, which was used for the manual transcription in the rare circumstances when the eye-tracker did not see the driver's eyes.

Method

We chose a within-subjects factorial design experiment with navigation type (*nav*) as our independent variable. We collected multiple dependent variables: PDT on the road ahead, number of collisions with other objects in the simulated world, NASA-TLX score, level of agreement with preferential statements (2 preferential statements using 5-point Likert scales) and average-based driving performance measures expressed through variances of lane position, steering wheel angle and velocity. In each case, higher values of driving performance measures indicate deterioration. We also calculated average velocity. All driving performance variables were obtained from the simulator at a frequency of 10 Hz, while the eye-tracker data was obtained at 60 Hz.

As shown in Figure 3.16, participants drove on two lane city roads which included ambient vehicles (about 6 vehicles per street segment), moving pedestrians, traffic signs and lane markings. Lanes were 3.6 meters wide. Participants were instructed to drive as they normally would in real life and to obey all traffic laws. They were also instructed (and trained) to pay attention to unexpected events, such as pedestrians emerging from behind parked vehicles (Figure 3.16) or vehicles braking suddenly. These unexpected events are not uncommon in city driving. Furthermore, the ability to avoid

collisions when such unexpected events occur is a valuable (although coarse) measure of driving performance.



Figure 3.16 Simulated two-lane city road with a pedestrian emerging from behind a parked vehicle.

For all three PNDs, the participants drove a different route with two unexpected events in each case. Figure 3.17 shows the whole navigation route (solid red line), road segments selected for the analysis (dashed red lines in the zoomed-in areas), locations of the unexpected events (numbers 1 and 2), start/end locations (green hexagons) and side streets (thin, solid blue lines). The first route included traveling from north to south. For the second route we reversed the direction of travel (south to north), while the third route was the mirror image of the first route. In short, all three routes were of the same length (about 10 km) and complexity. However, the turn-by-turn directions for each route were different. Thus, there was no risk of participants remembering navigation instructions from the previous route. All intersections along the charted route were either T or four-way intersections. This required the participants to listen (if they chose not to look at the screen) the whole voice direction in order to be able to decide which way to turn. Each route had both long (400 and 800 meters long) and short (200

meters long) segments with many intersections on the given path. On average it took about 15 minutes to traverse a route. The presentation order of the routes was the same for all participants. A total of 18 participants (average age 20.5) took part in the experiment.

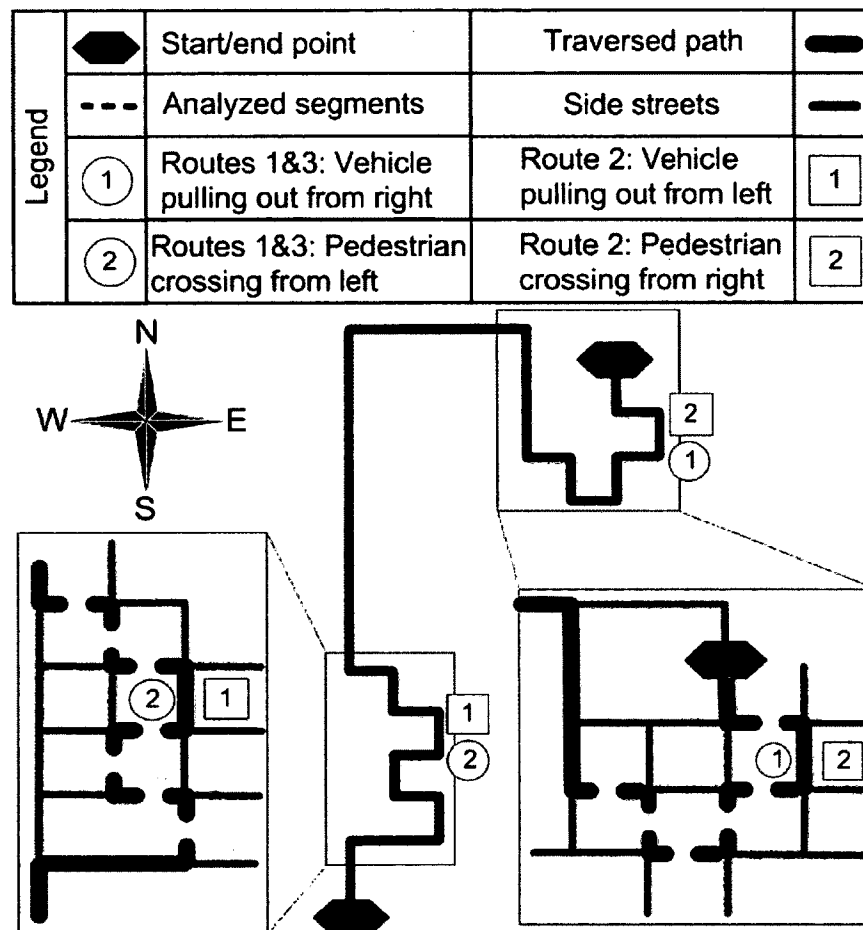


Figure 3.17 Simulated route with the segments selected for analysis.

The city routes in this experiment can be broken up into segments by treating roads between two intersections as separate segments. We calculated all of our driving performance (except the number of collisions) and visual attention results from 13 short segments (dashed red lines in Figure 3.17). All 13 short segments had the same characteristics, thereby controlling factors that could potentially confound our results. In

particular, the segments were 200 meters long measured from the centers of the adjacent intersections. The participants did not encounter any unexpected events (represented by 1 and 2 in Figure 3.17) in the 13 segments we used to analyze visual attention and driving performance. Unexpected events often require sudden braking and steering wheel motion, which in turn can result in very large first differences and variances for these measures, making comparisons with other segments difficult. For the purpose of counting the number of collisions only, we used 15 short segments, including the ones with unexpected events, since collisions are more likely to occur there.

In analyzing all of the segments, we excluded data collected over the first 60 meters and the final 40 meters of a segment, and analyzed data generated over $(200 - 60 - 40) = 100$ meters. This was done because driving performance tends to be different between the excluded and analyzed portions of the segments. For example, at the beginning of a segment, drivers are completing the turning maneuver that is necessary to get through the previous intersection. At the end of a segment, they are decelerating before entering the next intersection and possibly even approaching one of the sides of the lane depending on the direction of the upcoming turn. Thus, the resulting variances can be much larger than those encountered away from intersections, which makes it difficult to compare excluded and analyzed portions of segments.

After filling out the consent forms and personal information questionnaires, participants were given an overview of the driving simulator and descriptions of the three navigation devices. Next, they proceeded to complete three navigation experiments, one with each of the PNDs. Before each condition, we provided the participants with about 5 minutes of training using that PND. For training, users followed PND navigation

instructions in a city environment similar to the one experienced during the real experiment. In order to circumvent order effects, we counterbalanced the presentation order of the PNDs between participants.

General Results

Table 3.1 describes visual attention directed towards the road, LCD screen and the rest of the cabin. Note that the results presented here regarding visual attention may differ slightly from the results published in [36]. The reason is that the current results are obtained after applying the glance filtering procedure described in Section 3.1.2 (pg. 97), while the raw eye-tracker data was used in study [36]. Overall, the differences are very small and did not affect the outcomes of the statistical analyses.

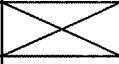
	AR	SV	SPND	p-value
road	96.48	86.69	89.38	<0.0001
LCD		9.77	7.19	<0.0001
cabin	3.27	3.01	2.96	0.798

Table 3.1 PDT on the road, LCD and the rest of the cabin as a function of PND type.

A repeated-measures ANOVA revealed a significant main effect of the navigation type on PDT on the road ($F(2,34)=83.789$, $p<0.0001$). Post-hoc comparisons indicated significant differences between all pairs: AR and SPND ($p<0.0001$), AR and SV ($p<0.0001$), and SV and SPND ($p=0.003$). We can see that the overall PDT on the road was 96.48%, 89.38% and 86.69% for AR, SV and standard PND, respectively. The difference of 9.79% between AR and SV indicates that on average for every minute of driving drivers spent about 5.87 seconds less looking at the road in case of SV PND. What is very interesting to note is that SV required even more visual attention than the

SPND. This was corroborated by a repeated-measures ANOVA with PND on the LCD as a dependent variable comparing SV and SPND in isolation. Again, a significant difference was detected ($F(1,17)=21.391$, $p<0.0001$). We also confirmed that PDT on the rest of the cabin was not significantly affected by the PND type ($p=0.798$).

To closer investigate the effects on visual attention, we calculated the number and duration of off-road glances for each PND. The left graph of Figure 3.18 shows the average glance duration, while the right graph shows the average number of glances. Since more than one glance may occur on each segment, we aggregated all off-road glances for each of the PNDs and performed statistical analysis using a one-way ANOVA. The analysis revealed a significant main effect ($F(2,759)=12.6036$, $p<0.0001$) of navigation type *Nav* on glance duration. Post-hoc comparisons indicated significant differences between all pairs: SPND and AR ($p<0.0001$), SV and AR ($p=0.0029$) and SPND and SV ($p=0.0324$). As we can see in the left graph of Figure 3.18, the average glance durations are 0.45, 0.58 and 0.53 seconds for AR, SPND and SV PND, respectively.

We applied the same procedure for the number of glances directed off-road as well, since we wanted to use the same statistical methods for the same family of variables. A one-way ANOVA indicated a significant main effect of navigation type ($F(2,690)=115.3878$, $p<0.0001$). Pairwise comparisons revealed significant differences between all pairs ($p<0.0001$). The right graph in Figure 3.18 shows the average number of glances to be 0.48 (AR), 1.22 (SPND) and 1.6 (SV).

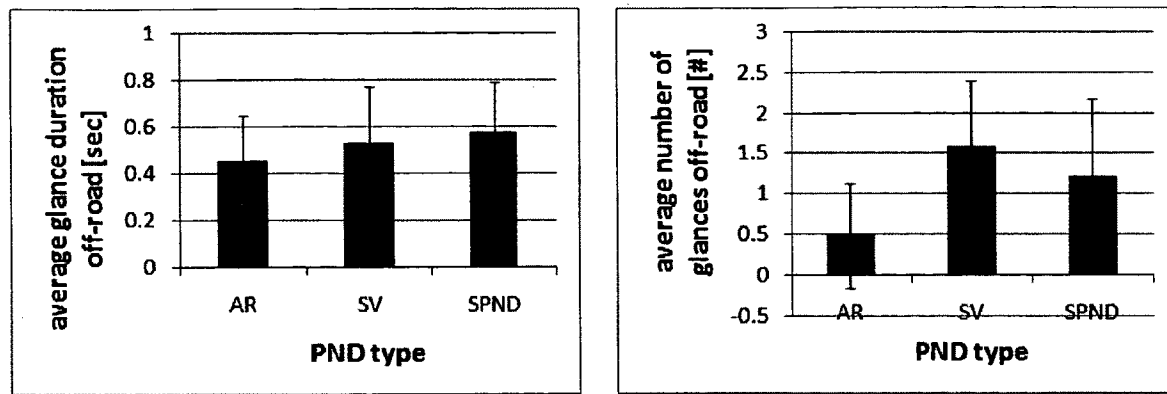


Figure 3.18 Average duration (left) and number (right) of glances directed off road for the three PNDs per segment.

Since the nature of the AR PND is such that the participants did not have to look away from the road to obtain navigation directions, segments without off-road glances often occurred in this condition. Even though ANOVA is robust to departures from normality (especially with large data samples), we intended to take a conservative approach and also conducted the analyses of the number of glances using a non-parametric Kruskal-Wallis test, which does not require the assumption of normal distribution. To be consistent, we also performed the non-parametric analysis in case of glance durations, as well. The results entirely match the ones obtained using a one-way ANOVA. Namely, significant main effects of the navigation type have been observed for both number of glances and glance duration: $\chi^2=197.7495$, $p<0.0001$ and $\chi^2=33.8531$, $p<0.0001$, respectively. If we look at pairwise comparisons (using a Wilcoxon Signed Ranks test), significant differences for number of glances have been observed between all pairs ($p<0.0001$). Similarly, in case of glance duration we observed significant differences between all pairs: SV and AR ($p=0.0003$), SPND and AR ($p<0.0001$) and SPND and SV ($p=0.002$).

Subjective estimates of cognitive load were estimated using NASA-TLX questionnaire. Users' average NASA-TLX ratings were 28.7, 38.7 and 33.4 for the AR, SV and SPND, respectively. We performed a one-way ANOVA to examine the effect of PND on these subjective workload ratings. Our analysis revealed a significant main effect of *Nav* on workload ($F(2,24)=6.759$, $p=0.005$). Post-hoc comparisons indicated that participants experienced significantly less load using the AR than the SV PND ($p<0.0001$). No difference was observed between SV and SPND ($p=0.136$). Even though the difference between AR and SPND ($p=0.097$) is not significant at the 0.05 level, it is significant at the 0.1 level, so we can conclude that a strong trend exists.

Using 5-point Likert scales participants indicated their level of agreement (highly agree, agree, undecided, disagree, highly disagree) with two preferential statements presented in Table 3.2. The numbers shown under AR, SV and SPND columns specify the percent of participants who highly agreed/agreed (white cells) or highly disagreed/disagreed (shaded cells) with each statement. Note that the percentages do not always sum up to 100% since some participants were undecided. For each statement we performed a Friedman non-parametric test with respect to *Nav*.

Statement	Agreement	AR [%]	SV [%]	SPND [%]	p (χ^2)
My driving was best when using [AR/SV/SPND] interface.	highly agree or agree	72.2	11.1	38.9	0.014 (8.49)
	highly disagree or disagree	16.7	61.1	50	
I prefer to have a [AR/SV/SPND] for navigation.	highly agree or agree	66.7	22.2	38.9	0.023 (7.53)
	highly disagree or disagree	16.7	72.2	27.8	

Table 3.2 Level of agreement with two preferential statements.

Table 3.2 shows a significant main effect of *Nav* on the subjective judgment about best driving performance ($p=0.014$). Participants ranked AR PND very highly (72% highly agreed or agreed) in comparison to others, while both SV and SPND were perceived as detrimental to driving (61% and 50% disagreed or highly disagreed, respectively). Using the Wilcoxon Signed Ranks test for pairwise comparisons, we found significant differences between AR and SPND ($p=0.027$) and AR and SV ($p=0.003$). Clearly, most participants felt that the AR PND allowed for the best driving performance.

A significant main effect of the navigation type on the subjective preference for a particular PND was detected ($p=0.023$). Responses to this preferential statement in Table 3.2 indicate that participants liked the AR PND. Using the Wilcoxon Signed Rank test, we found that participants significantly preferred the AR PND to both the SV ($p=0.007$) and SPND ($p=0.045$) and that participants significantly preferred the SPND over the SV PND ($p=0.038$).

Based on the visual attention results, we can say that, as expected, HUD-based AR PND allowed users to keep their eyes on the road more than the HDD-based SPND and SV PND. This result was also supported by the NASA-TLX scores which showed that participants found the SV PND more difficult to use than the AR PND. The fact that we observed a difference in PDT between SV and SPND suggests that PDT is not solely a function of display modality. Rather, it is likely that participants found it difficult to resolve differences between the real world and SV images. This explanation is supported by the significantly more frequent glances ($p<0.0001$) at the LCD display in the SV condition than with the SPND (1.25 and 0.87 glances on average, respectively). Subjective assessments also support this explanation.

There were no collisions with pedestrians or ambient traffic for any PND on segments without unexpected events. There were 8 collisions in total with vehicles on segments with unexpected events: 2 for AR, 3 for SV and 3 for SPND. Clearly, the occurrence of collisions did not depend on the PND type.

Despite all of the observed differences in visual attention and subjective assessments, we found no significant differences between the three PNDs regarding any of the average-based driving performance measures (in all cases $p > 0.05$). Figure 3.19 shows the average variances calculated for lane position (upper left), steering wheel angle (upper right) and velocity (bottom left) as well as the average velocity (bottom right). This suggests that any distractions by these PNDs were not high enough to be detected using long-term averages of driving performance measures.

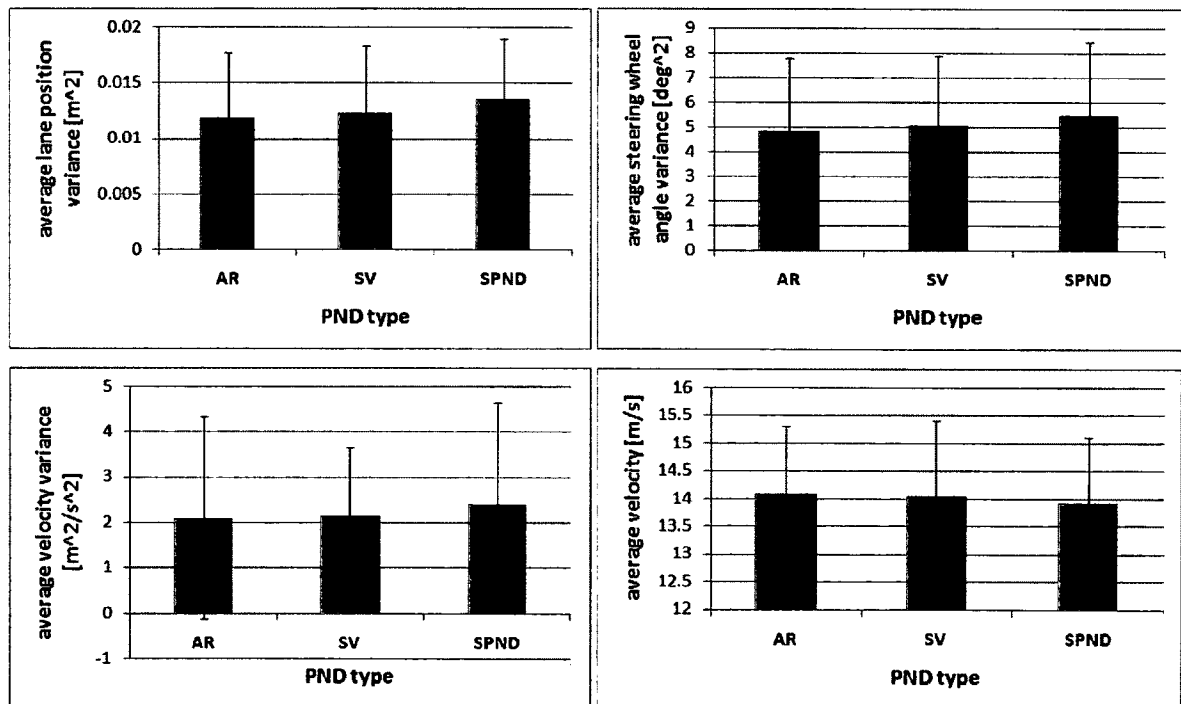


Figure 3.19 Average variances of lane position (upper left), steering wheel angle (upper right) and (bottom left) and average velocity (bottom right).

Cross-Correlation Results

Figure 3.20 shows the cumulative cross-correlation functions (obtained using the Equation 3.2) for all three navigation devices when the steering wheel angle was used for calculating the driving performance sequence (DPS).

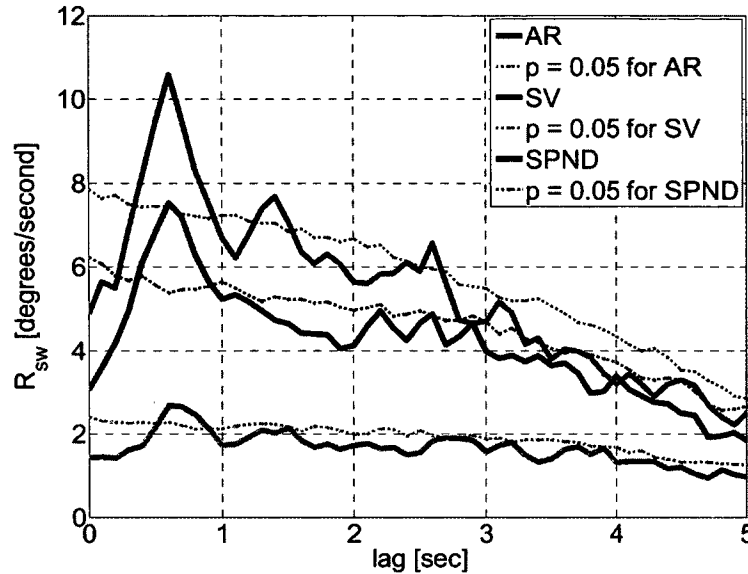


Figure 3.20 Cumulative steering wheel angle cross-correlation functions calculated for AR, SV and SPND.

Similarly, Figure 3.21 depicts the same cumulative cross-correlation functions except that the DPS sequence was obtained using the lane position. In both figures, solid lines represent cross-correlation functions, while the dotted lines indicate the significance levels of $p=0.05$ obtained using the randomization method described in detail in section 3.1.4 (pg. 113). The significance level of 0.05 is commonly used among researchers and it is applied for other analyses in this dissertation. Therefore, we decided to apply the same significance level in all of the following figures for the purpose of establishing the significance of the cross-correlation peaks. A separate significance level was calculated for each device.

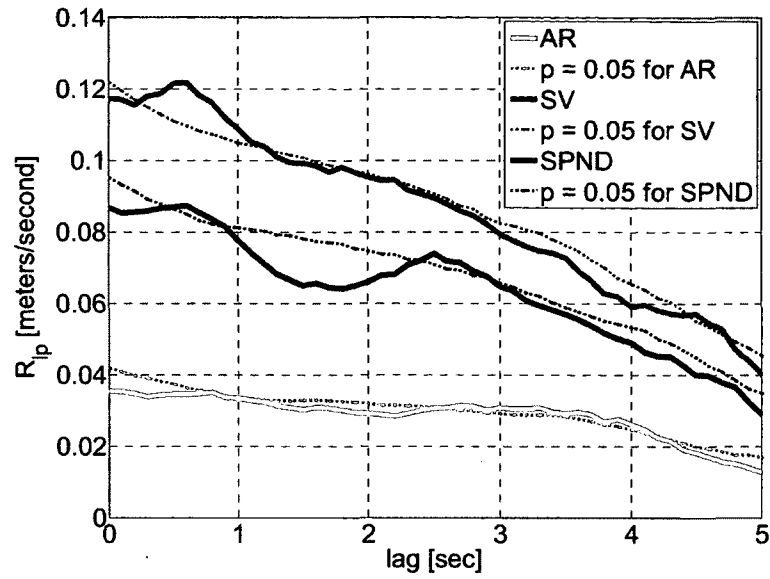


Figure 3.21 Cumulative lane position cross-correlation functions calculated for AR, SV and SPND.

Before we continue with the analysis of the results, we have to make two notes. First, in this experiment the overall segment durations were relatively short (about 7 seconds), relative to the maximum evaluated lag of 5 seconds. Thus, the calculated cross-correlation functions have the observed tendency to decrease with lag due to the decreasing overlap between the EGS and DPS sequences (see Equation 3.2). And second, even though the graphs presented in Figure 3.20 and Figure 3.21 have the same shapes as the ones published in [36], the orders of the magnitudes differ. The reason for this is a different normalization scheme which was applied in [36]: the cross-correlation result was normalized by the number of samples of each experimental segment which was used in the calculations (about 70 samples/segment). Furthermore, in this dissertation we use the AVC function to transform each DPS sequence. This produces the *true* absolute first difference as opposed to just absolute first difference used in [36]. Since we employed

the sampling interval of 0.1 second, it introduces a normalization factor of 10 ($1/0.1$) compared to absolute first difference (see Equation 3.1). Therefore, the overall normalization factor which accounts for the difference between the results presented here and in [36] is about 700.

As we can see in Figure 3.20 there are statistically significant peaks in $R_{sw}[lag]$ for all three PNDs, at the $p=0.05$ level. The most prominent peaks appear at the lag of 0.6 sec for all three PNDs. These peaks indicate that on average, the periods of looking away from the road ahead are followed by a larger change in the steering wheel angle (possible corrective actions) than in usual circumstances. Note that there is also a significant peak for AR. Even though in case of AR navigation the participants did not have a specific device to look at, the occurrence of this peak is sound, since the participants cast occasional glances towards the speedometer, steering wheel or dashboard. Nevertheless, its magnitude is much smaller compared to SV and SPND. Similarly, Figure 3.21 shows the most prominent peaks for $R_{lp}[lag]$ at the lag of 0.6 sec for SV and SPND and at 0.8 sec for AR.

Significant peaks that occur far from the edge of the glance (located at the lag of 0 seconds) are due to the nature of the cross-correlation formula where glances separated in time may get correlated with each other's effects on driving performance (as described in Section 3.1.1, pg. 111). There are two very good examples at the lags of 1.4 and 2.6 seconds in the case of R_{sw} for SV PND. If we take into account that the average separation between falling edges of glances for SV PND is 0.97 seconds and using the finding that the largest changes occur on average 0.6 seconds after the glance, we would expect that the average separation between one glance and a change in driving

performance coming from another glance should be about $0.97 + 0.6 = 1.57$ seconds. This matches very well with the distant cross-correlation peak observed at 1.4 seconds. The lag at 2.6 seconds is even further away from the edge of the glance and it is a very long time interval while driving, during which accidents may occur unless the driver timely applies a necessary action. Specifically, at the speed of 35 MPH, which was the posted speed limit in this experiment, the car would have travelled 40.67 meters in 2.6 seconds. Therefore, any necessary action will likely be applied sooner. Additionally, both of these peaks have much smaller magnitudes compared to the highest peak at 0.6 seconds.

In the previous paragraphs we showed that each of the most prominent cross-correlation peaks are significant at the level of $p=0.05$. However, another question that can be asked here is whether the peaks among different PNDs are significantly different as well. Since in this case we have individual cross-correlation functions for each navigation device, we conducted statistical comparisons using the two approaches presented in Section 3.1.5 (pg. 118). Table 3.3 shows the results of the analysis. There are two main columns in the table: the left column shows the results obtained by comparing the most prominent cross-correlation peaks only, while the right column compares the areas under the curves for a range of lags (specifically, from 0 to 1 second). Also, two main rows indicate the specific cumulative cross-correlation functions that are being compared: steering wheel angle or lane position. Bolded values indicate significant differences at the specified level. There was a significant main effect ($p<0.001$) of the navigation type (*Nav*) for each approach, thus allowing us to perform pairwise comparisons. Similarly, pairwise comparisons also revealed significant differences

($p < 0.001$) between all three PNDs. If we look at the results of both approaches in concert, we can conclude that the effect of using the three PNDs exists not only where the most prominent peaks occur, but also over the range of lags surrounding the peaks.

Cumulative steering wheel angle cross-correlation	Comparing Highest Peaks			Comparing Areas Below Curves		
	Main effect of <i>Nav</i>		p < 0.001	Main effect of <i>Nav</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	SV-AR	SV-SPND	AR-SPND	SV-AR	SV-SPND	AR-SPND
	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Cumulative lane position cross-correlation	Main effect of <i>Nav</i>		p < 0.001	Main effect of <i>Nav</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	SV-AR	SV-SPND	AR-SPND	SV-AR	SV-SPND	AR-SPND
	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001

Table 3.3 Statistical comparisons between cumulative cross-correlation results for three PNDs.

Since the significant differences have been observed between the magnitudes of the most prominent cross-correlation peaks, we can now rank the size of the effect for the three PNDs. For example, in the case of cross-correlations calculated for steering wheel angle, we can see that an average cumulative effect of glances directed off-road contributes to an absolute change (AVC) on the steering wheel amounting to 10.65, 7.53 and 2.682 degrees/second for SV, SPND and AR, respectively. If we use AR as the reference, we can see that the cumulative effect of looking away from the road in case of SV is $10.65/2.682 = 3.97$ times higher relative to AR. Similarly, the cumulative effect in case of SPND is $7.53/2.682 = 2.81$ times higher relative to AR PND.

Both Figure 3.20 and Figure 3.21 demonstrate that the effect size is the smallest for the AR and the largest for the SV. The relatively large difference in effect size between AR, on the one hand, and SV and SPND on the other might be attributed to the difference in display type: HUD for AR vs. HDD for SV and SPND. However, we also see that the cross-correlation peaks for SPND are consistently smaller than for SV. This indicates that resolving differences between SV images and the observed world may be cognitively taxing (certainly time consuming), even more so than receiving directions from a 2D map. Note that our simulated world and SV images were very similar, as the season, the weather and the time of day were identical in the two simulations that generated these images. In the real-life scenarios these variables are likely to be different between the outside world and street view data. Thus, the observed separation in the cross-correlation results may be emphasized even further in real-life conditions.

If we look at the way cross-correlation function is calculated (Equation 3.2), it represents a combination of both driving performance and visual attention measures. Since both of these measures are manifestations of cognitive load, it is of interest to observe how the cross-correlation results compare to other estimates of cognitive load. The cross-correlation function defined in Equation 3.2 provides a cumulative effect of interacting with the three PNDs. Thus, it would be the most appropriate to perform the comparison with another measure that provides an overall estimate of cognitive load. It is for this reason that we decided to use the results obtained from the NASA-TLX questionnaire. Figure 3.22 shows strong positive relationships between prominent peaks in both cumulative R_{sw} and R_{lp} and NASA-TLX results.

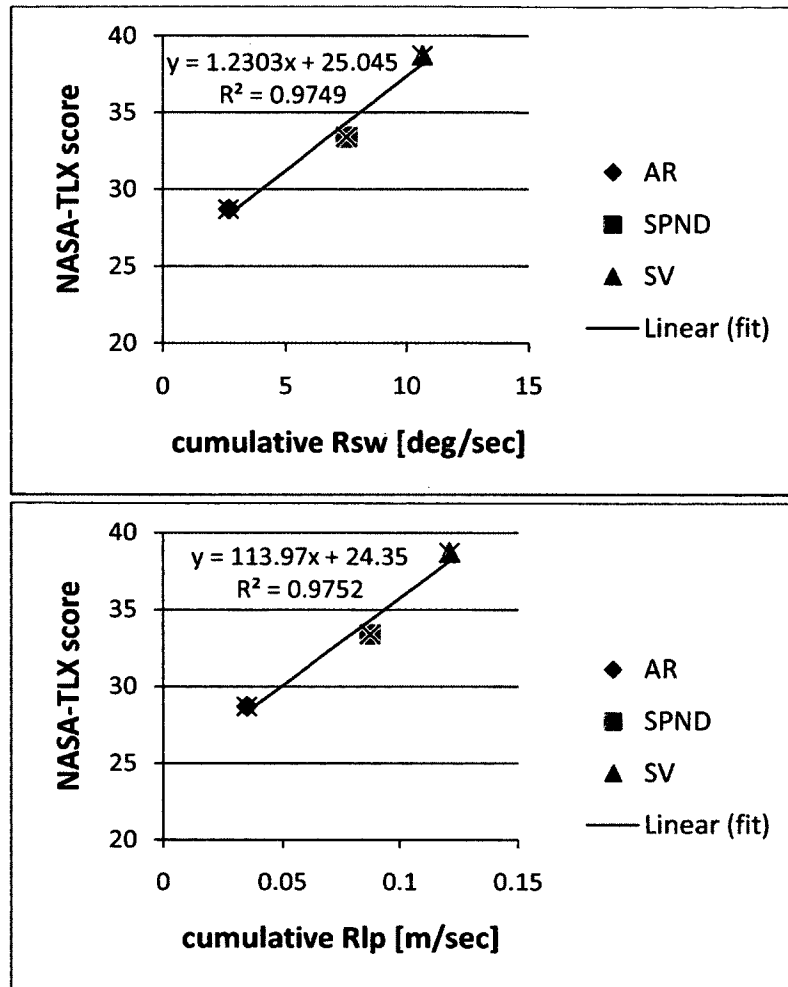


Figure 3.22 Magnitudes of prominent peaks in cumulative R_{sw} (upper graph) and R_{lp} (lower graph) vs. NASA-TLX score for AR, SV and SPND.

A simple linear fit in both cases revealed very high coefficients of determination ($R^2 > 0.97$). This is an important result, since it indicates that both the cumulative cross-correlation peaks and the subjective estimates of cognitive load point to the same conclusion regarding the three PNDs: AR is perceived as the one with the smallest impact on cognitive load, followed by SPND and SV PNDs.

In Section 3.1.1 (pg. 109) we also presented a modified cross-correlation definition (Equation 3.4) which allows us to estimate the average change (AVC) in driving performance measures after looking away from the road *per individual glance*.

Figure 3.23 and Figure 3.24 show per-glance cross-correlation functions obtained for the three PNDs for steering wheel angle and lane position, respectively. The results observed in these figures are very important, because they indicate that significant effects of individual glances directed off-road exist besides the cumulative effects.

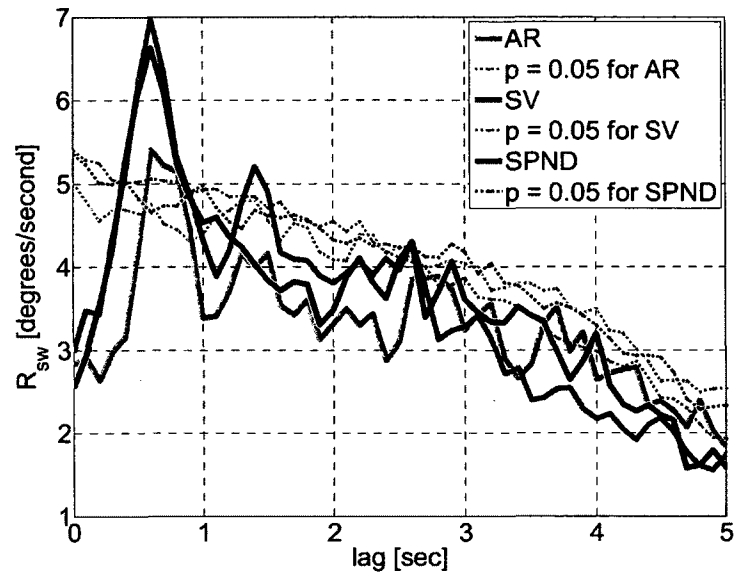


Figure 3.23 Per-glance steering wheel angle cross-correlation functions calculated for AR, SV and SPND.

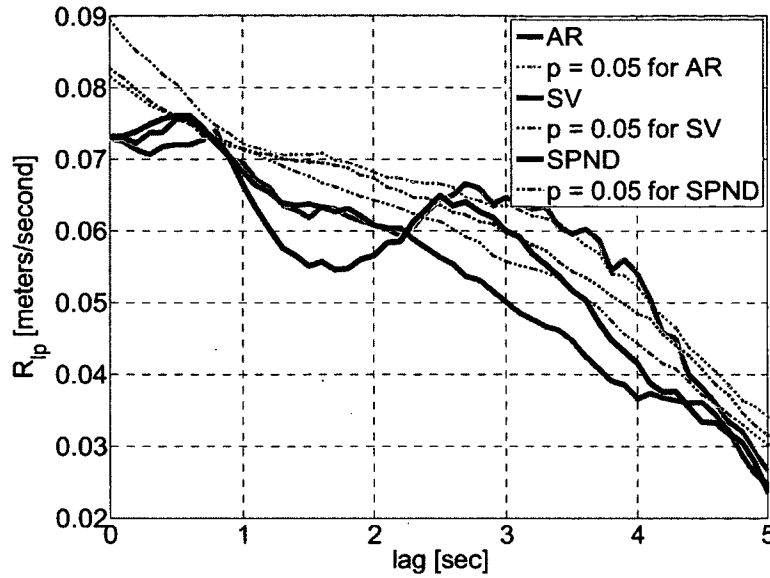


Figure 3.24 Per-glance lane position cross-correlation functions calculated for AR, SV and SPND.

In Figure 3.23 we can see that significant per-glance steering wheel angle cross-correlation peaks exist for all three PNDs at the lag of 0.6 seconds. Similarly, Figure 3.24 shows significant per-glance lane position cross-correlation peaks at the same lag for SV and SPND. No significant peak is detected for AR PND (although the most prominent peak at 0.8 seconds is just below the significance level). Even though visual attention to the road was very high for AR PND (96.48% PDT to the road ahead), the influences of individual glances were still detected by the per-glance steering wheel angle cross-correlation (notice that the highest peak is above the significance level). However, this was not the case for per-glance lane position cross-correlation. This difference in the observed effect can be attributed to the difference in dynamics that exists between lane position and steering wheel angle: faster dynamics in case of steering wheel angle and slower in case of lane position. This is very obvious if we look at example amplitude spectra of both variables based on the real data from this study. Figure 3.25 shows the

raw data, while Figure 3.26 shows the amplitude spectra for lane position and steering wheel angle for one example experimental segment.

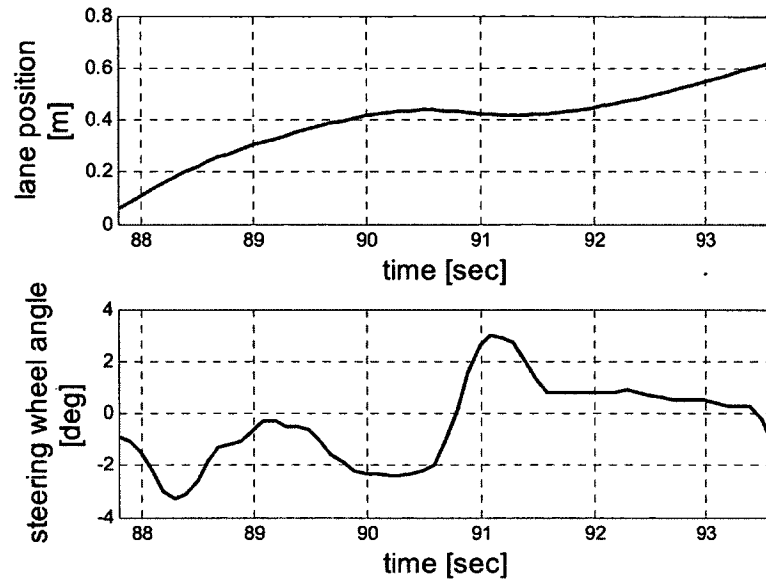


Figure 3.25 Example raw data for lane position and steering wheel angle.

We can see that lane position is dominated by low frequencies ($f < 0.5$ Hz), while steering wheel angle has a considerable frequency content beyond 0.5 Hz as well.

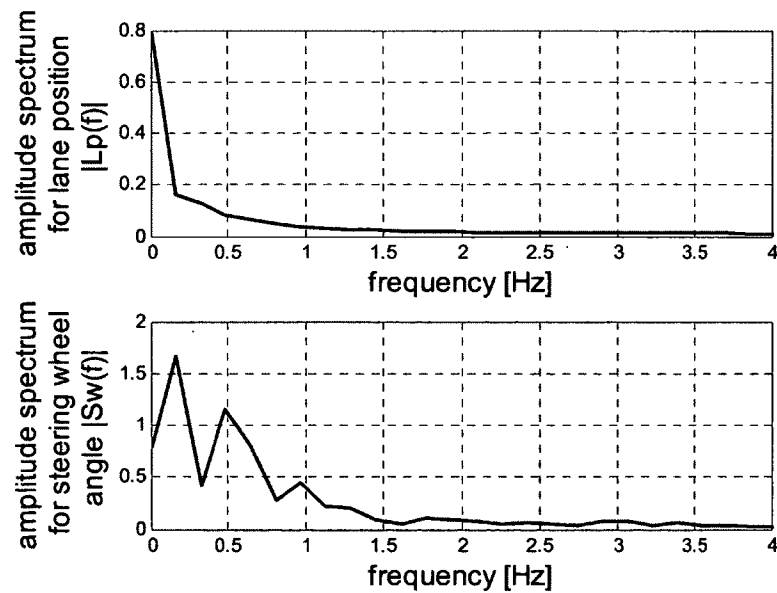


Figure 3.26 Amplitude spectra for example lane position and steering wheel angle data.

Table 3.4 shows the statistical comparisons between the per-glance cross-correlations for the three PNDs.

Per-glance steering wheel angle cross-correlation	Comparing Highest Peaks			Comparing Areas Below Curves		
	Main effect of <i>Nav</i>		p = 0.0421	Main effect of <i>Nav</i>		p = 0.0043
	Pairwise comparisons			Pairwise comparisons		
	SV-AR	SV-SPND	AR-SPND	SV-AR	SV-SPND	AR-SPND
	p = 0.0126	p = 0.2304	p = 0.1376	p = 0.0013	p = 0.3816	p = 0.0101
Per-glance lane position cross-correlation	Main effect of <i>Nav</i>		p = 0.8663	Main effect of <i>Nav</i>		p = 0.8241
	Pairwise comparisons			Pairwise comparisons		
	SV-AR	SV-SPND	AR-SPND	SV-AR	SV-SPND	AR-SPND
	N/A	N/A	N/A	N/A	N/A	N/A

Table 3.4 Statistical comparisons between per-glance cross-correlation results for three PNDs.

We can see that a significant main effect of *Nav* is detected ($p=0.0485$) in case of per-glance steering wheel angle cross-correlation when comparing the magnitudes of the most prominent peaks. After performing pairwise comparisons a significant difference is detected between SV and AR ($p=0.0126$), although the difference between AR and SPND ($p=0.1376$) is close to the significance level of 0.1. The area below the curves approach also detected a significant main effect of the navigation type ($p=0.0043$), while the pairwise comparisons detected differences between SV and AR ($p=0.0013$) and AR and SPND ($p=0.0101$).

Neither approach indicated the existence of the main effect of navigation type in case of per-glance lane position cross-correlation results, which is not surprising given the large overlap between the curves that can be seen in Figure 3.24. However, it is still

important to notice that significant peaks exist in case of SV and SPND, which indicates that the effects of individual glances directed off-road do exist, even though they are not different between the two PNDs.

There are two important findings resulting from the per-glance steering wheel angle cross-correlation results. First, we can see that a significant peak exists for all three PNDs indicating that significant effects of individual glances directed off-road exist in each case. The largest influence occurs right after the gaze moves back to the forward road, even though this effect was not obvious through average-based measures. And second, the observed differences between SV and AR and SPND and AR indicate that average glances directed off-road in case of SV and SPND influence driving and cognitive load more compared to AR PND.

The lack of significant difference between SV and SPND in case of per-glance steering wheel angle cross-correlation is not that surprising given the large overlap that can be seen between the two in Figure 3.23. This result indicates that SV and SPND are not very different when observed from the standpoint of an individual glance (instance of interaction). However, we can argue that more instances of interaction (glances directed off-road) in case of SV (1.6 glances) compared to SPND (1.22 glances) contributed to the observed difference in the cumulative results.

3.2.2 Highway Driving and iPod Interactions

In order to investigate how the situation would change for different driving environment and interaction modality, we conducted a study with another popular in-vehicle device: the iPod. One reason we selected this particular device is that some

negative effects of using an MP3 player on driving have been documented in the research literature. For example, Salvucci et al. [18] looked at driving performance degradation while choosing music, podcasts and videos on a fifth generation iPod. The study found that selecting media while driving significantly affected both lateral and speed deviation. We expected to observe similar results in our study as well. However, if participants interact with the MP3 player infrequently over the course of an experiment, and/or if the individual interactions are short, based on the previous experience, we expected that the negative influence on driving performance might be difficult to demonstrate by observing average-based driving performance measures.

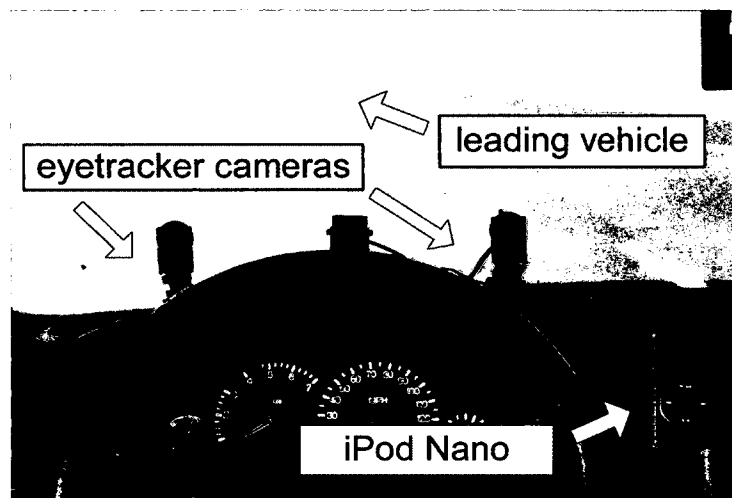


Figure 3.27 Experimental setup inside the simulator cabin.

We used an iPod Nano device as our MP3 player. As shown in Figure 3.27 the iPod was attached to a board on the right side of the steering wheel. We decided to place the iPod in a fixed location, so that all drivers would experience the same experimental setup. This location allowed for very easy manual access and required a small change in eye gaze direction away from the roadway, compared to when the player is held in the hand or placed anywhere else on the central console. For example, Salvucci et al. [18]

located their iPod in a holder mounted fairly low on the central console. Thus, it is possible that the effects of iPod interactions we are about to present here would have been even larger had we decided to use the same location. The iPod was also connected to the simulator's speakers, so the drivers were able to hear the songs they were asked to play. A total of 12 participants (average age 21.5) participated in the experiment.

Method

As a primary task, the participants were instructed to follow a yellow lead vehicle travelling at 55mph (88.5km/h) (see Figure 3.27). The road was a straight portion of a divided highway with three lanes in each direction, each 3.6 meters wide. Both the lead and the participant vehicle were travelling in the middle lane. Roads were presented in daylight with light (approximately 1 vehicle every 2 seconds), random ambient traffic in the other two lanes. Participants were instructed to drive as they normally would and to obey all traffic laws.

In addition to the primary task, the participants experienced three conditions describing their engagement in the secondary task of interaction with the iPod:

1. *No secondary task - baseline (B)*. In this condition participants did not have any additional task. Their only concern was to follow the lead vehicle while driving safely.
2. *Easy iPod interaction (E)*. In this condition the participants were given a number of simple operations to perform on the iPod. These operations included: selecting a menu option, playing the previous, current and next song, pausing, increasing/decreasing volume and fast-forwarding a song (Figure 3.28). All participants completed the same 10 operations in the same order. Individual

interactions were initiated automatically by custom software. Every 40 seconds the participant heard a voice prompt by the computer to perform an interaction. We decided to initiate interactions every 40 seconds in order to allow enough time for participants to complete the previous interaction, as well as enough down-time between individual interactions. As we can see in Figure 3.28, most actions required simple clicks on one of 5 available buttons. Increasing/decreasing volume and rewinding/advancing a song also required short scrolling.

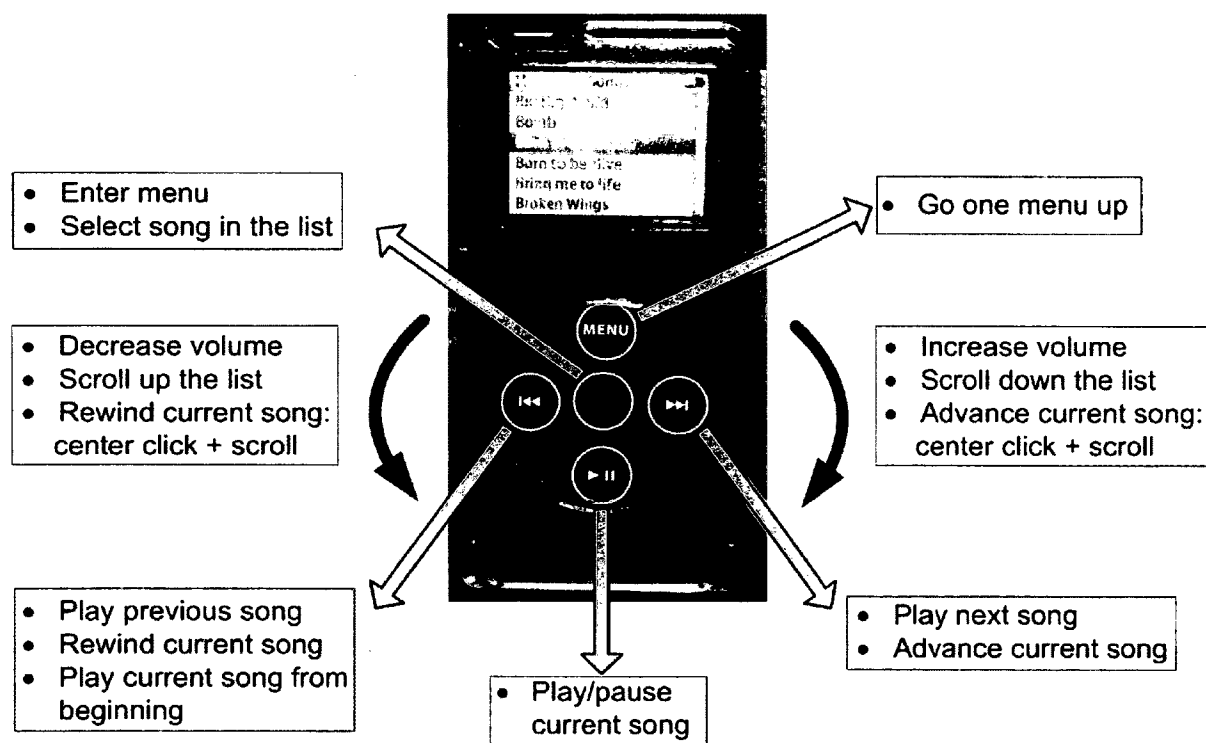


Figure 3.28 iPod interactions participants performed during the experiment.

3. *Difficult iPod interaction (D).* Under this condition participants were given the name of a song (by a computer voice) which they needed to locate in the list of all songs preloaded on the device and play it. The iPod contained a total of 347 songs, which were sorted alphabetically by title. They had to search for 10 songs during the experiment. These songs were given to the participants in alphabetical

order, so that they would need to scroll only in one direction to find the next one. This simplified the task somewhat, since changing the direction of the search is much more challenging. The titles of the sought songs were distributed approximately uniformly throughout the list, so the participants would experience the same level of difficulty when searching for each song. Specifically, they had to scroll on average 36 songs from the current one in order to find the next song (range of scrolling = [33; 42], SD = 2.71). If the rotation is performed relatively slowly (for example, one 360° turn per 1 second), one 360° turn moves the selection pointer about 16 songs (this is important to notice, since faster rotation exponentially increases scrolling speed). The names of the songs as well as their order were the same for all participants. The interaction timing followed the same pattern as for the easy task: a new task was initiated every 40 seconds.

We conducted a within-subjects factorial design experiment with the interaction type as our primary independent variable, *Int*. The levels of this variable were: no secondary task - baseline (B), easy iPod task (E) and difficult iPod task (D). The order of *Int* was counterbalanced among the participants. We measured the following dependent variables: PDT on the forward road, average glance duration, average number of glances, average driving performance measures expressed through the variances of lane position, steering wheel angle and velocity, average velocity and subjective estimates of cognitive load based on NASA-TLX score.

Our experiment presented participants with straight highway routes. We broke up the routes into segments by treating parts of the highway where participants engaged in the secondary task as separate segments. Since there were 10 interactions in total (for E

and D conditions) and for each interaction the participants had a maximum of 40 seconds, we calculated all of our dependent variables using data from those 10 segments. Even though the baseline (B) condition did not employ any interactions, the segmentation was possible by dividing the experiment into ten, 40-second-long segments. This allowed for direct comparison between the three conditions. All segments had the same characteristics, thereby controlling factors that could potentially confound our results. In particular, the segments were relatively long, at about 924 meters.

General Results

To assess the effect of secondary task complexity on visual attention, we performed a repeated measures one-way ANOVA using PDT on the forward road as the dependent variable. As expected, we found a highly significant main effect ($F(2,22)=108.991$, $p<0.0001$) (see Figure 3.29). In addition, all the post-hoc pair-wise comparisons between baseline, easy and difficult conditions were also highly significant (for all pairings $p<0.0001$). The average values of PDT for the three levels of *Int* showed large differences: B – 94.62%, E – 85.12% and D – 72.98%. For the difficult task this would amount to spending 16.21 seconds of every minute not looking at the road ahead. The same measure for the easy task would be 8.93 seconds, while for the baseline it would amount only to 3.23 seconds of inattention to the roadway for each minute of driving. These results show that the drivers kept their visual attention focused significantly more on the inside of the car (and away from the roadway) as the secondary task got more complex. This can be explained with the fact that more difficult iPod tasks demanded more visual attention.

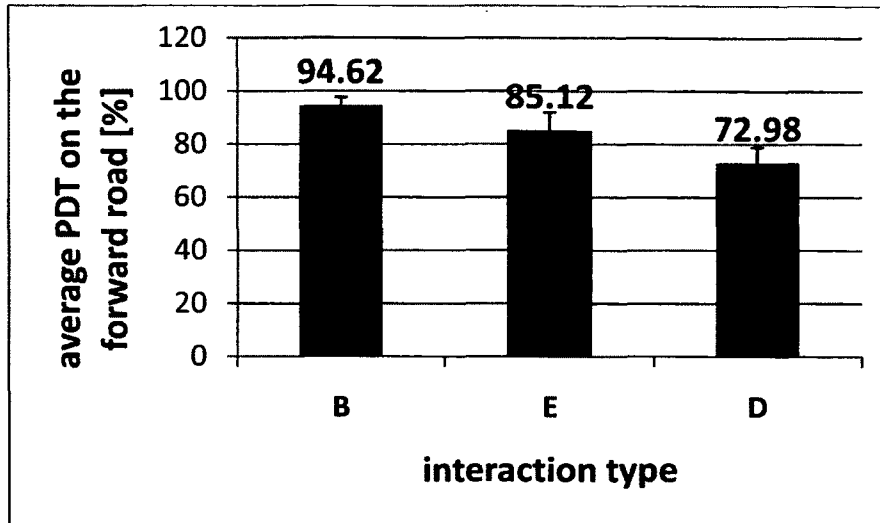


Figure 3.29 Average PDT on the forward road.

To obtain more fine-grained information pertaining to visual attention, we calculated the average duration and number of glances directed away from the road for each interaction type. As in the previous study, we aggregated all glances directed off-road for each of the three interaction types, since multiple glances can occur on each experimental segment. The left graph of Figure 3.30 shows the average glance durations to be 0.59, 0.79 and 0.98 seconds for B, E and D condition, respectively. A one-way ANOVA indicated a significant main effect of the interaction type *Int* on glance duration ($F(2,2533)=100.5490$, $p<0.0001$). Post-hoc comparisons revealed significant differences between all possible pairs at $p<0.0001$. To keep the analysis procedure consistent with the previous study, we also conducted a non-parametric Kruskal-Wallis test. The analysis indicated the same conclusions: a significant main effect of the interaction type ($\chi^2=200.6734$, $p<0.0001$) and significant differences between all pairs ($p<0.0001$).

The right graph of Figure 3.30 shows that the average number of glances directed away from the road is 3.49, 7.24 and 10.78 for B, E and D condition, respectively. As with the duration of glances, a one-way ANOVA revealed a significant

main effect of the interaction type on the number of glances ($F(2,352)=149.6802$, $p<0.0001$). Post-hoc pairwise comparisons indicated significant differences between all possible pairs: D and B ($p<0.0001$), D and E ($p<0.0001$) and E and B ($p<0.0001$). Non-parametric Kruskal-Wallis test revealed the same conclusions: significant main effect of the interaction type ($\chi^2=178.5063$, $p<0.0001$) and significant differences between all interaction pairs ($p<0.0001$).

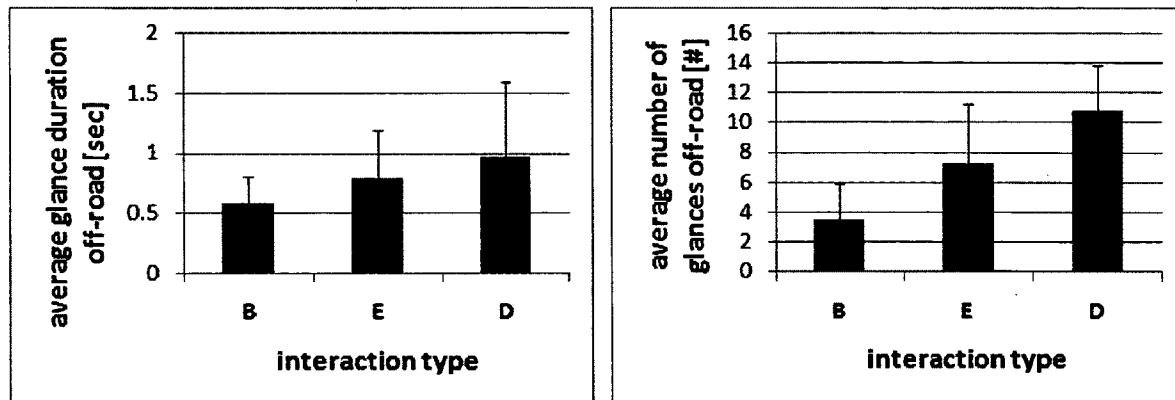


Figure 3.30 Average duration (left) and number (right) of glances directed off road.

Regarding subjective estimates of cognitive load, a repeated measures ANOVA revealed a significant main effect of interaction type on the NASA-TLX score ($F(2,22)=10.977$, $p<0.0001$) (Figure 3.31). Post-hoc pairwise comparisons indicated significant differences between baseline and difficult ($p=0.001$) and baseline and easy ($p=0.013$) conditions. No significant difference has been observed between easy and difficult conditions ($p=0.075$). However, it can be considered marginally significant, since it is close to the significance level of 0.05 and lower than 0.1. If we take into account that the NASA-TLX score for D (41.58) is larger than for E (32.31), we can conclude that a strong trend towards D being more cognitively loading does exist.

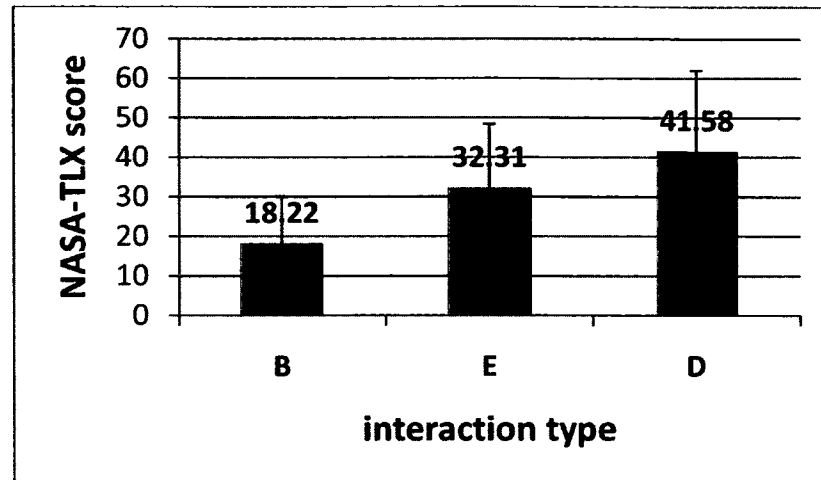


Figure 3.31 Average NASA-TLX score.

We performed a repeated measures one-way ANOVA for each of the driving performance measures with *Int* as the independent variable. Figure 3.32 shows the average variances for lane position (upper left), steering wheel angle (upper right), velocity (bottom left) and average velocity (bottom right).

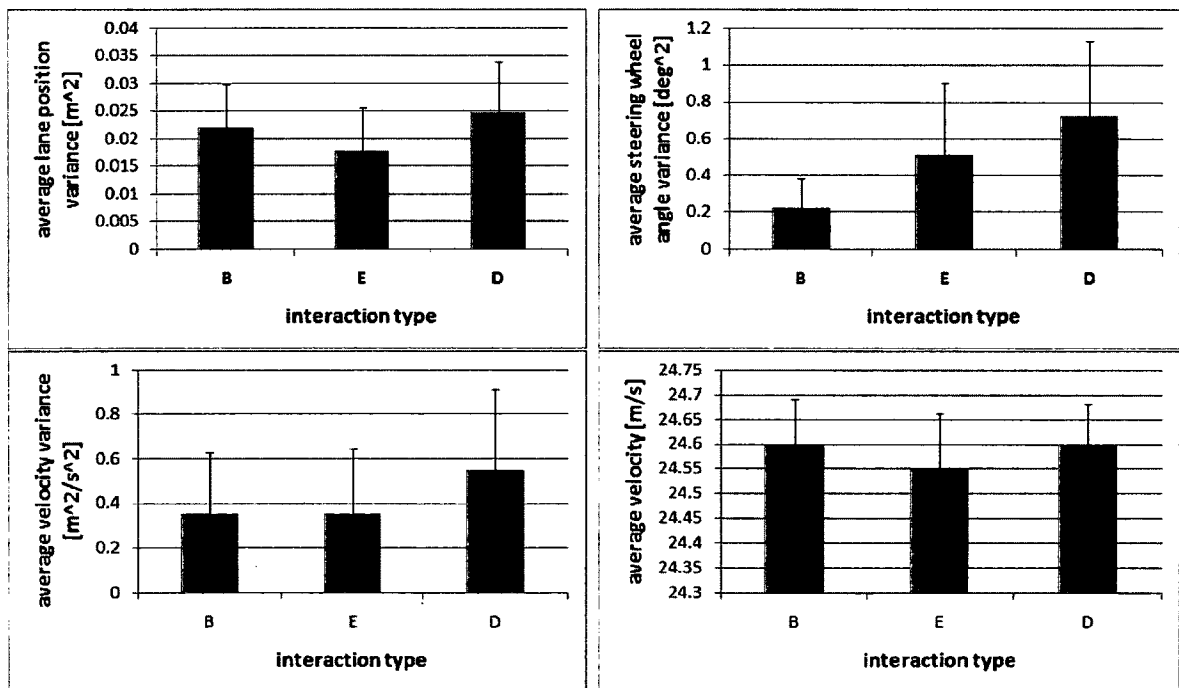


Figure 3.32 Average variances of lane position (upper left), steering wheel angle (upper right) and velocity (bottom left) and average velocity (bottom right).

Table 3.5 outlines the results of the statistical analyses for all dependent driving variables. For each variable we present the corresponding F- and p-values for the main effect, as well as the pairwise comparisons conditional on the significance of the main effect. Note that bolded p-values indicate significance at the 0.05 level.

Dependent variable	F-value	p-value	p-values for pairwise comparisons		
			B - E	B - D	E - D
Lane position variance	$F(2,22) = 4.031$	0.032	0.093	0.380	0.004
Steering wheel angle variance	$F(2,22) = 11.835$	<0.0001	0.026	<0.0001	0.052
Velocity variance	$F(2,22) = 3.709$	0.041	0.990	0.073	0.048
Average velocity	$F(2,22) = 2.265$	0.127	N/A	N/A	N/A

Table 3.5 Results of statistical analyses for all dependent driving variables.

As we can see from Table 3.5 there is a significant main effect of the interaction type on variances of lane position, steering wheel angle and velocity, but not on average velocity. The results are mixed when it comes to pairwise comparisons:

1. Lane position variance - the only significant difference was observed between E and D conditions ($p=0.004$). The comparison of B and E revealed a weakly significant difference ($p=0.093 < 0.1$), which indicates potential trends.
2. Steering wheel angle variance - pair-wise comparisons revealed significant differences between B and E ($p=0.026$) and B and D task conditions ($p<0.0001$). The difference between E and D conditions is just over the significance level of 0.05 ($p=0.052$), therefore, it can be considered marginally significant.

3. Velocity variance – similar to the variance of lane position, the only significant difference was detected between E and D conditions ($p=0.048$). The difference between B and D is approaching significance ($p=0.073$), which indicates existing trends.

The lack of significant differences between some of the conditions for all dependent variables is surprising, given that we observed a very significant impact of the interaction type on all aspects of visual attention as well as the subjective estimates of cognitive load. This is another example that the lack of sensitivity of the average-based driving performance measures can occur with manual-visual interactions as well.

Another interesting result is that the variance of lane position in B condition is larger than the variance in E condition. It is possible that the participants paid less attention to the car's position in B compared to E condition due to the uneventful nature of the B task (unencumbered driving, just following the lead vehicle). Even though B is not significantly different from other conditions, this can create a misleading impression about the ranking of the experimental conditions.

Cross-Correlation Results

Figure 3.33 and Figure 3.34 show cumulative cross-correlation functions calculated for all three interaction conditions using steering wheel angle and lane position as DPS sequences, respectively. Solid lines represent cross-correlation functions, while dotted lines represent their corresponding significance levels of $p=0.05$. We have to note here that, unlike the previous study (“Exploring Augmented Reality Navigation Aids”), the segment durations in the current study (and the iPod study which will be presented in Chapter 4) were relatively long (about 40 seconds) relative to the maximum evaluated lag

of 5 seconds. Thus, the decreasing overlap between the EGS and DPS sequences is not a significant factor in the computation of cross-correlations (see Equation 3.2). As a result, we have “flatter” appearances of the results in case of iPod studies.

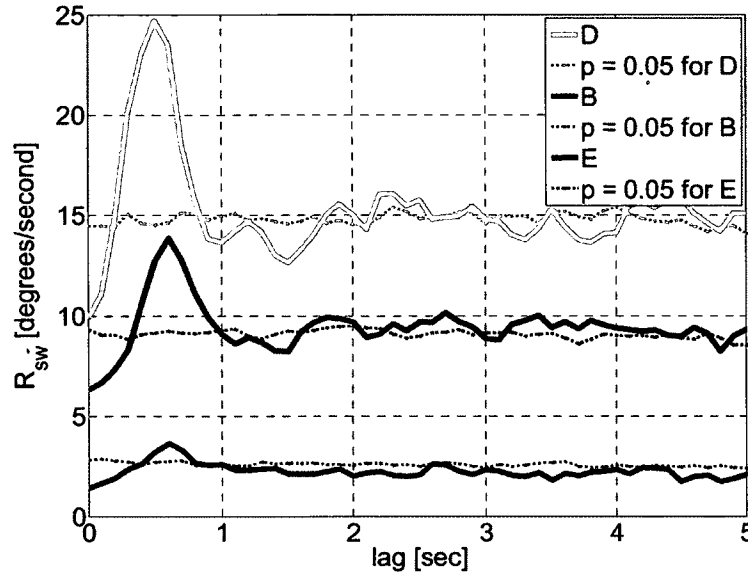


Figure 3.33 Cumulative steering wheel angle cross-correlation functions calculated for D, B and E conditions.

As we can see in Figure 3.33 statistically significant peaks in $R_{sw}[lag]$ exist for all three interaction types. Each peak represents the average cumulative amount of angular change on the steering wheel over the course of interaction with the iPod. The existence of these prominent peaks indicates that on average there is a pronounced absolute change (AVC) in steering wheel angle about half a second after returning the gaze to the forward road. The most prominent peaks appear at the lags of 0.5 seconds for D, 0.6 seconds for E and 0.6 seconds for B condition. It is no surprise that the peak exists even in the B condition (even though it is fairly small compared to others), since the participants cast occasional glances towards the speedometer, steering wheel or

dashboard, and some of those glances might have resulted in larger changes once the gaze returned to the road.

Similarly, Figure 3.34 shows the statistically significant peaks in $R_{lp}[lag]$ at the lag of 0.6 sec for D, 1 sec for E (although, there is an almost entirely flat area in the cross-correlation function between 0.7 and 1 second) and 0.8 sec for B condition. Even though the peak in the B condition is very small, for both steering wheel angle and lane position it indicates that even during unencumbered driving glances directed away from the road have small but significant cumulative effects. However, these effects are many times smaller compared to other conditions, suggesting that the B condition cumulatively provides the smallest impact on driving and cognitive load.

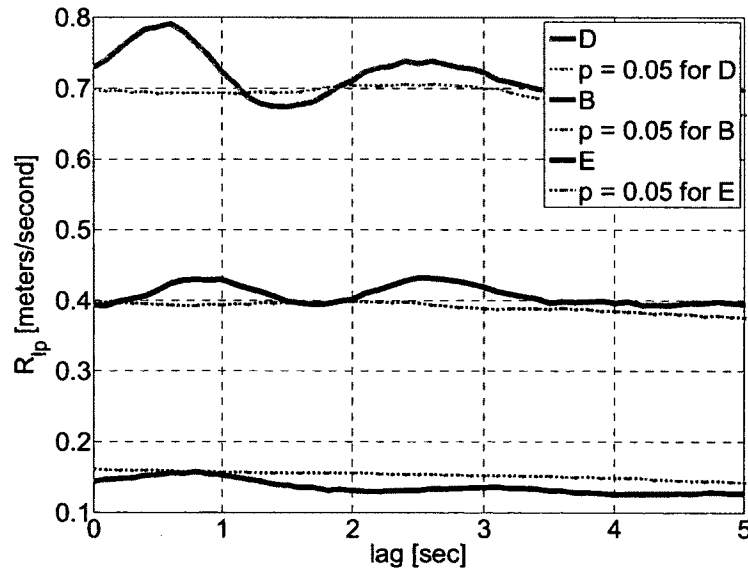


Figure 3.34 Cumulative lane position cross-correlation functions calculated for D, B and E conditions.

Similar to the previous study, some distant cross-correlation peaks are visible in case of D and E conditions. Since the segments in this study are long (40 seconds) and

the nature of the secondary task is such that the participants can interact with the iPod for potentially long periods of time, glances directed off-road can be very dispersed. Thus, it can be expected that distant cross-correlation peaks may occur at various lags. Regarding cumulative steering wheel angle cross-correlation functions the distant peaks are both far from the edge of the glance (>1.5 seconds) and considerably smaller in magnitude compared to the most prominent peaks. The distant peaks that we can see in cumulative lane position cross-correlation functions (see Figure 3.34) occur more than 2 seconds away from the edge of the glance. During a 2 second time interval the vehicle travels 49.16 meters at the posted speed limit of 55 MPH in this experiment. Hence, if a reaction to an unexpected event (such as the lead vehicle braking or approaching the edge of the road) is necessary after returning the gaze to the road, it is likely that it would be applied earlier.

So far we have demonstrated that the significant cumulative effect of looking away from the road exists for both difficult and easy interaction with the iPod. The effect is visible even in the baseline condition in case of cumulative steering wheel angle and lane position cross-correlation functions. What we intend to explore now is whether these individual effects are different from each other and how they rank. For this purpose we use the comparison procedure presented in Section 3.1.5 (pg. 118).

Table 3.6 outlines the results of the statistical comparisons. As we can see, both approaches detected a significant main effect ($p < 0.001$) of interaction type *Int* for both R_{sw} and R_{lp} . Pairwise comparisons revealed significant differences ($p < 0.001$) between all possible pairs of interactions. Based on these results we can rank the three interaction types with respect to the average cumulative effects they produce on steering

wheel angle and lane position over the course of interaction with the iPod: D has the largest influence, followed by E and B. By comparing the magnitudes of the most prominent peaks, we can also determine the relative differences between the individual conditions. For instance, the magnitudes of the most prominent peaks for the cumulative steering wheel angle cross-correlation functions are as follows: 24.7, 13.88 and 3.647 degrees/second. If we use B as a reference, we can see that D produces $24.7/3.647 = 6.77$ times larger effect than the B condition. Equivalently, E produces $13.88/3.647 = 3.81$ times larger effect compared to B. If we compare D and E conditions alone, we can see that D results in $24.7/13.88 = 1.78$ times stronger cumulative effect on steering wheel angle compared to E.

Cumulative steering wheel angle cross-correlation	Comparing Highest Peaks			Comparing Areas Below Curves		
	Main effect of <i>Int</i>		p < 0.001	Main effect of <i>Int</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
Cumulative lane position cross-correlation	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
	Main effect of <i>Int</i>		p < 0.001	Main effect of <i>Int</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001

Table 3.6 Statistical comparisons between cumulative cross-correlation results for three interaction types.

This ranking matches our initial expectations and can be explained as follows. Just driving and following a lead vehicle on a straight highway with light ambient traffic is likely to be fairly simple (B condition). On the other hand, interactions with the iPod

can introduce varying levels of difficulty. Even though easy interactions (E condition) typically involved simple button presses, they still resulted in significant cumulative effects on steering wheel angle and lane position. This reflects the fact that the participants had to divide their mental and visual attention between the driving and the iPod task. Furthermore, since the interaction is manual-visual, the participants also had to remove one hand from the wheel, which introduced the physical distraction as well. This agrees with the predictions of the Wickens' multiple resources theory [10], since many of the resources are shared between the driving and the interaction task. All of these effects can be expected only to increase in case of difficult iPod interactions (D condition). Namely, even though we intended to help our participants by issuing the sought songs in the alphabetical order, the task was still fairly demanding since it involved actively scanning the contents of the list. This placed a high burden on the participants in all processing stages: visual, mental and manual response. Thus, it is not surprising to see the D condition to produce the largest cumulative effect on both driving measures.

The explanations from the previous paragraph are also supported by the subjective estimates of cognitive load obtained through the NASA-TLX score (see Figure 3.31). Similar to the previous study, we wanted to observe how the cross-correlation results compare to subjective measures. Figure 3.35 shows positive relationships between the magnitudes of the most prominent peaks in R_{sw} and R_{lp} and NASA-TLX results. We can see that in both cases the coefficients of determination are very high ($R^2 > 0.96$), which indicate that the cross-correlation peaks offer the same conclusion as the subjective estimates about cognitive load changes: D has the highest impact on cognitive load followed by E and B conditions.

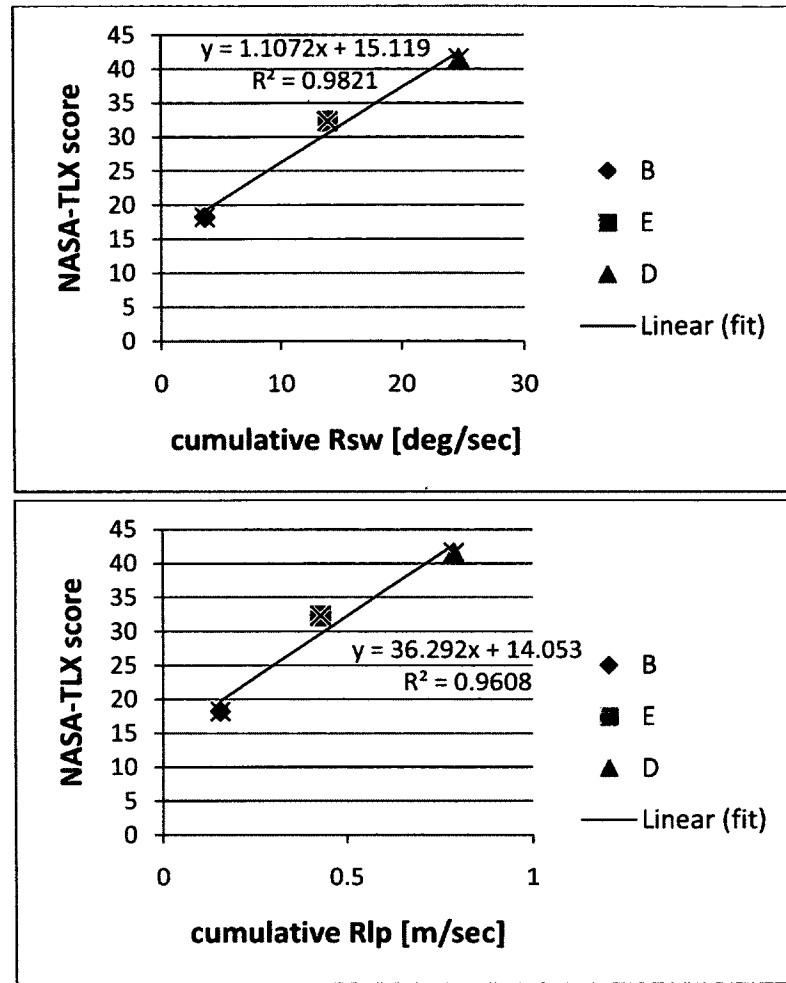


Figure 3.35 Magnitudes of prominent peaks in cumulative R_{sw} (upper graph) and R_{lp} (lower graph) vs. NASA-TLX score for B, E and D conditions.

Figure 3.36 and Figure 3.37 show per-glance cross-correlation functions calculated for all three interaction types using the steering wheel angle and lane position as DPS sequences, respectively. Again, solid lines represent cross-correlation functions, while dotted lines represent their corresponding significance levels of $p=0.05$.

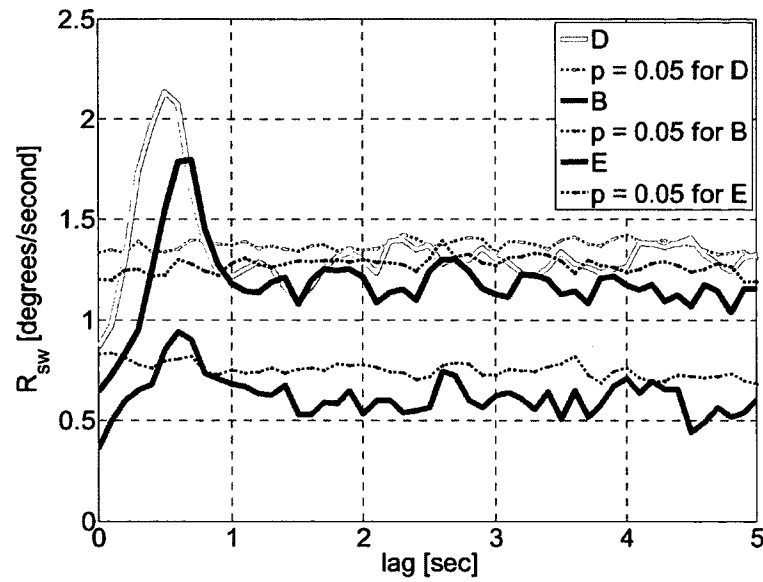


Figure 3.36 Per-glance steering wheel angle cross-correlation functions calculated for *D*, *B* and *E* conditions.

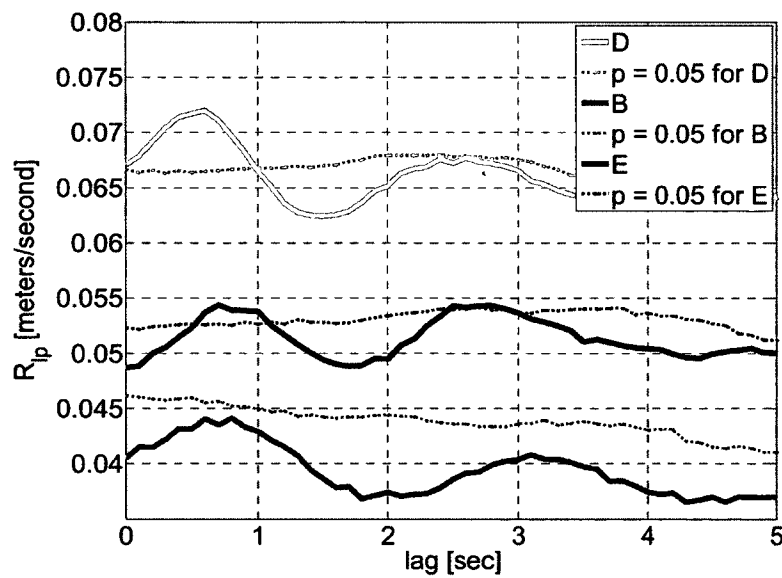


Figure 3.37 Per-glance lane position cross-correlation functions calculated for *D*, *B* and *E* conditions.

As we can see in Figure 3.36, statistically significant peaks in R_{sw} exist for all three interaction types. These significant peaks indicate the average amount of angular change on the steering wheel contributed by an *average glance* directed off-road. The most prominent peaks appear at the lags of 0.5 seconds for D, 0.7 seconds for E and 0.6 seconds for B condition. As with the cumulative response, the existence of the peak in the B condition can be explained by the participants' occasional glances towards the speedometer, steering wheel and dashboard. The highest peak is in the case of D (2.136 deg/sec), followed by E (1.798 deg/sec) and B (0.941 deg/sec) conditions.

Figure 3.37 shows that significant peaks exist in R_{lp} in case of D and E conditions, but not B condition. The most prominent peaks appear at 0.6 seconds for D and 0.7 seconds for E condition. The magnitudes of the highest peaks for D and E conditions indicate that an average glance contributes to an absolute change (AVC) in lane position equaling to 0.072 and 0.054 meters/second, respectively. Even though it is not significant, the most prominent peak for the B condition is located at 0.8 seconds. Since this peak is lower than the significance level, it indicates that an individual glance on average does not result in significant changes in lane position after returning the gaze to the road. Even though a significant peak was observed for the B condition in case of steering wheel angle, it is possible that the changes were not influential enough to result in significant effects on lane position. As suggested in the previous study, the difference in dynamics between steering wheel angle and lane position is one possible explanation for the observed result. Figure 3.38 and Figure 3.39 illustrate this assertion by presenting the raw data and the amplitude spectra for steering wheel angle and lane position for one example segment, respectively.

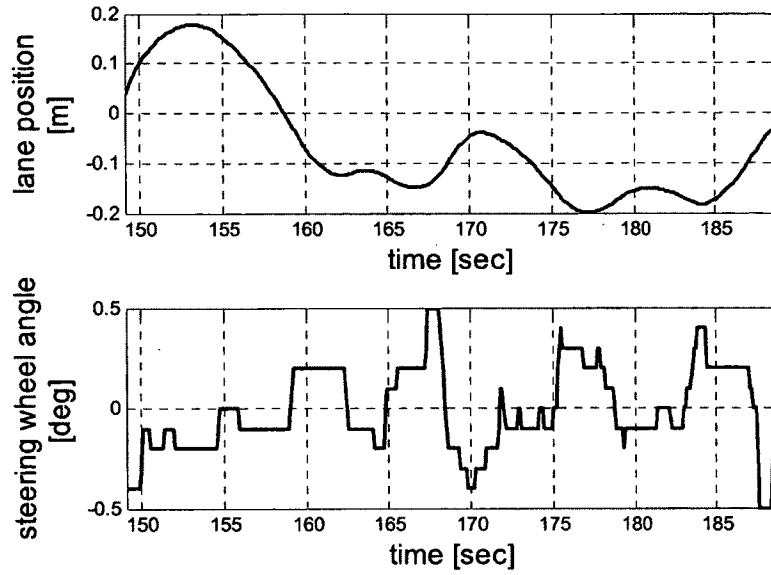


Figure 3.38 Example raw data for lane position and steering wheel angle.

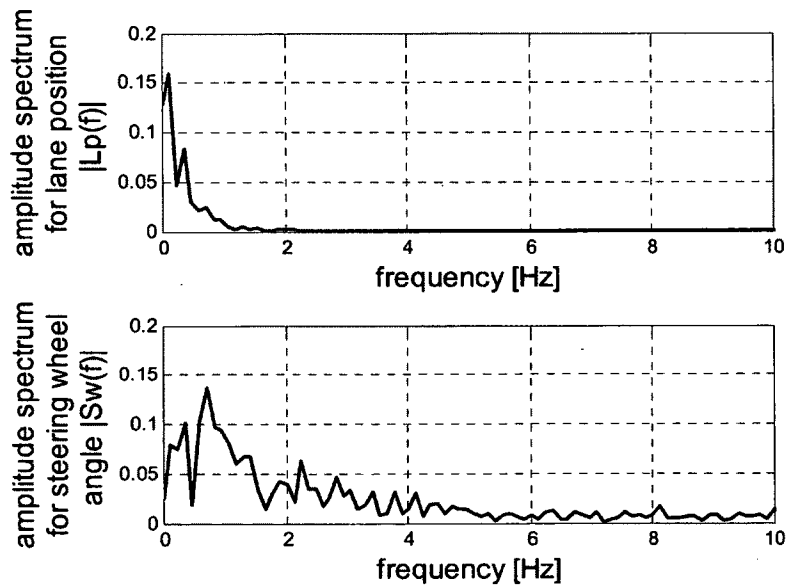


Figure 3.39 Amplitude spectra for example lane position and steering wheel angle data.

We can see that the frequency content of lane position practically dies out after 1 Hz. On the other hand, there is a considerable frequency content in case of steering wheel angle for frequencies larger than 1 Hz as well. The findings regarding the B

condition are valuable, since they indicate that under unencumbered conditions (no secondary task) the drivers were able to pay sufficient attention to their speed (since they were instructed to follow a lead vehicle at a constant distance) and maintain good driving performance.

Table 3.7 shows the results of the statistical comparisons between per-glance cross-correlation functions for all three conditions. As we can see, significant main effects of the navigation type ($p < 0.001$) exist for both R_{lp} and R_{sw} . Post-hoc pairwise comparisons using both procedures (highest peaks and areas below the curves) revealed significant differences ($p < 0.05$) for all possible pairs in case of both lane position and steering wheel angle cross-correlation results. Based on this if we look at the per-glance cross-correlation functions for lane position, we can say that the highest influence exists in the case of D, followed by E and B conditions. Per-glance cross-correlation functions for steering wheel angle provided the same ranking.

Per-glance steering wheel angle cross-correlation	Comparing Highest Peaks			Comparing Areas Below Curves		
	Main effect of <i>Int</i>		p < 0.001	Main effect of <i>Int</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
	p < 0.001	p < 0.001	p = 0.0046	p < 0.001	p < 0.001	p = 0.0018
Per-glance lane position cross-correlation	Main effect of <i>Int</i>		p < 0.001	Main effect of <i>Int</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
	p < 0.001	p = 0.007	p < 0.001	p < 0.001	p = 0.0065	p < 0.001

Table 3.7 Statistical comparisons between per-glance cross-correlation functions for three interaction types.

3.2.3 General Discussion of the Results

Addressing Hypotheses

The purpose of this chapter was to address the following hypotheses:

- H1 – which proposed an initiator-based approach to quantifying cumulative effects of secondary task engagements,
- H2 – which proposed an initiator-based approach to quantifying the effects of individual instances of secondary task engagements,
- H_{RP} – which proposed two ways of ranking the above cumulative and instance-based effects of secondary task engagements.

The introductory sections of this chapter provided detailed explanations of the procedures proposed in the above hypotheses. Both H1 and H2 are initiator-based, which means that they account for the potential causes of changes in driving performance measures of interest. Specifically, we expected that changes (possible corrective actions) in driving performance measures (such as lane position and steering wheel angle) may occur following glances directed off-road. We used two driving simulator studies which employed multimodal interactions with in-vehicle devices (three PNDs and an iPod) in testing these hypotheses. Both types of devices result in visual attention being directed away from the forward road, since interactions with PNDs require visual while interactions with iPod require manual-visual modalities. Therefore, we can conclude that glances directed off-road describe these interactions well and this is the reason why we decided to use those as “initiators” in our method.

As we had a chance to see in Sections 3.2.1 and 3.2.2, the effects of off-road glances were successfully detected through statistically significant cumulative and per-glance cross-correlation peaks.

In both studies (“Exploring Augmented Reality Navigation Aids” and “Highway Driving and iPod Interactions”) we observed significant cumulative cross-correlation peaks for both steering wheel angle and lane position. The fact that the most prominent peaks are statistically significant indicates that the corrective actions on average follow the reductions in visual attention to the road and that they do not occur by chance. Furthermore, since the method accounts for all off-road glances in concert, these peaks resemble the overall effects on driving and cognitive load produced over the course of interaction with these in-vehicle devices. These results indicate that H1 is supported.

Similarly, in both studies we observed significant per-glance cross-correlation peaks for all conditions, except for per-glance lane position cross-correlations for AR PND in the navigation study and B condition in the iPod study. These results are very important, because they indicate that the effects on driving and cognitive load exist not only when looked at from the cumulative standpoint (which resembles both the individual interactions and the frequency of those interactions), but also at the level of an average instance of interaction, that is average off-road glance in our case. The finding that significant per-glance lane position cross-correlation peaks were not observed in case of AR PND in the navigation study and B condition in the iPod study is valuable as well, because it suggests that the off-road glances under those conditions did not negatively affect driving and cognitive load. Therefore, we can conclude that H2 is also supported.

Using the procedure proposed in hypothesis H_{RP} we were able to rank the experimental conditions in both studies based on their cumulative cross-correlation results. This was in contrast to average-based measures which often did not provide enough sensitivity to distinguish between different conditions. For example, in the navigation study, no differences between PNDs have been observed using any of the average-based measures. Conversely, cumulative steering wheel angle and lane position cross-correlation results indicated differences between all three PNDs, with SV producing the largest impact on driving followed by SPND and AR PNDs. In the iPod study, the best sensitivity to different iPod tasks regarding average-based measures was obtained using the variance of steering wheel angle, which detected differences between all pairs. Other measures (variances of lane position and velocity) only detected differences between D and E conditions. On the other hand, cumulative steering wheel angle and lane position cross-correlation results detected differences between all pairs of interactions, with D resulting in the largest impact on driving followed by E and B conditions.

Similarly, we used the same procedure outlined in hypothesis H_{RP} for ranking the per-glance cross-correlation results. In the navigation study we detected differences in the per-glance steering wheel angle cross-correlation results between AR and SV and between AR and SPND. The ranking obtained based on these significant differences indicated that the individual glances directed off-road produced significantly smaller impact for AR as compared to SV and SPND. On the other hand, SV and SPND produced similar effects per average glance. In the iPod study we detected significant differences in per-glance steering wheel angle and lane position cross-correlation results

between all interaction types and the ranking matched the one obtained with the cumulative results: D resulted in the largest impact, followed by E and B conditions. The importance of these results cannot be emphasized enough, because they indicate that the majority of the tested experimental conditions differ even at the level of average glances directed off-road, which is a clear indicator that differences in cognitive load introduced by these different interaction types do exist. This conclusion provides an important insight into the type of interaction performed and can be used for comparing different designs. Based on the rankings obtained for both cumulative and per-glance cross-correlation results we can conclude that H_{RP} is supported.

Comparing Cross-Correlation and Average-based Results

The main advantage of our method that we set out to demonstrate is the ability to detect short-lived and/or infrequent deteriorations in driving performance that may easily be lost when analyzed using average-based measures. As we had a chance to see in this chapter, the results of two driving simulator studies clearly show that this is the case. These studies provided examples of multimodal interactions with in-vehicle devices which result in both short and long-lived effects on driving performance. Gazing towards the displays of PNDs as well as the short and simple manual-visual interactions with an iPod (E condition) are the examples of short-lived effects on driving. As we had a chance to see our method successfully detected the influences of these interactions through statistically significant cross-correlation peaks (both cumulative and per-glance). Furthermore, we detected significant differences between the majority of experimental conditions for both studies using our cumulative and per-glance cross-correlation results, even when those differences were not obvious using the average-based measures. This in

turn allowed ranking the tested conditions with respect to their influence on driving and cognitive load. Based on these results we can say that our method provides a very sensitive measure.

We have to note here that our cumulative measure is similar in nature to the average-based measures, since they both provide a high-level description of the experimental condition of interest. However, even when no other reference is available for comparison (i.e., only one experimental condition is being analyzed) our method provides more information compared to the average-based measures. Namely, average-based measures provide only one numerical quantity which describes the experimental condition of interest and unless a reference is available, we cannot draw any conclusions from it. On the other hand, our cumulative cross-correlation measure describes how the performance measures of interest change over time, when the largest change (most prominent peak) occurs and whether the change is statistically significant or not. This way we can determine whether the selected “initiators” actually have the suspected impact on driving and cognitive load.

If we look at the per-glance cross-correlation results, they provide even more information since they also allow observing the effect of an average instance of secondary task engagement, which cannot be obtained using the average-based measures.

Construct Validity

Our results suggest that construct validity of our proposed method regarding cognitive load estimation is supported. Namely, in both studies we obtained very strong positive relationships ($R^2 > 0.96$) between the most prominent peaks in cumulative steering wheel angle and lane position cross-correlation functions and the subjective

estimates of cognitive load obtained using the NASA-TLX questionnaire (see Figure 3.22 and Figure 3.35). It is very interesting to notice that even a simple linear function provided such a strong fit. Nevertheless, the shape of this relationship should be examined further in the future studies. The existence of this strong relationship is a very important finding, because it confirms that both measures indicate changes in cognitive load in the same direction. This means that our method provides another objective measure which may help in avoiding circular arguments, as suggested by Wickens [10]. Furthermore, in the iPod study the ranking of experimental conditions obtained through the variances of steering wheel angle matched the ranking obtained using both cumulative and per-glance steering wheel angle cross-correlation results.

Taking all of the above into account we can formulate three general conclusions:

- a) if the average-based measures provide enough sensitivity, then they provide the same conclusions as our cross-correlation measures,
- b) our cross-correlation measures complement the average-based measures when those do not provide enough sensitivity,
- c) in each of the above cases our instance-based (per-glance in our case) cross-correlation measures provide low level insight into individual instances of interaction which cannot be obtained using the average-based measures.

General Observations

Ranking of the cross-correlation results does not allow us to draw immediate conclusions about how using the different PNDs or interaction types with an iPod relate

to the risk of a collision. In fact, there were no collisions on any of the experimental segments used for data analysis in our studies which resulted from using the tested devices. The most reliable risk estimation is obtained from naturalistic driving studies resulting in large databases of real-life driving data. Various conditions have to align for the accidents to actually occur (recall our discussion of the Swiss cheese model of incidence occurrence presented in Chapter 2). Those can be identified through naturalistic studies, since they provide realistic context to the overall driving experience [129]. From those studies we know that accidents are very often preceded by driver distractions of various kinds. The distractions often result in deteriorations in driving performance. Thus, being able to judge the amount of deterioration that a particular interaction can produce is valuable and may suggest likely risk increases. This is exactly what we are seeing with our cross-correlation results, even though it is often not detected through average-based measures.

As we had a chance to see almost all of the observed most prominent peaks in the cross-correlation functions occurred around 0.6 seconds after returning the gaze to the road. We hypothesize that this observed lag may be related to the urgency to respond to the situation on the road ahead and the reaction time. How urgent the response should be depends on many factors, such as the lateral distance from the edge of the lane (the response may be faster if the vehicle is closer to the edge) and the existence of an obstacle. According to the literature review created by Kosinski [130], mean reaction time for college-age individuals (which agrees with the age group of our participants) is about 190 msec to detect visual stimulus. This can be compared with obtaining visual information about the position of the vehicle in the lane after returning the gaze to the

road ahead. Since the participants have at least one hand on the steering wheel throughout the drive, as soon as the visual stimulus is detected, the reaction can be applied. This agrees with our findings. Namely, the fact that we observed the largest change on average about 0.6 seconds after the gaze returns to the road indicates that the participants actually started applying the correction on the steering wheel earlier. It is also interesting to notice that the time when the largest change on the steering wheel occurs is very similar to the brake reaction time of 0.7 seconds observed in the literature for fully aware individuals [131]. These results provide insights into the potential sources of the behavior of the lag. However, further studies are required to investigate whether and how the lag varies depending on the characteristics of the driving and secondary tasks. Chapter 5 proposes multiple experimental settings which may help in achieving this goal.

One aspect that is worth discussing is the difference in shape between the lane position and steering wheel angle cross-correlation results. On average we can say that these measures are mirrored and provide the same conclusions regarding detection and ranking of secondary task engagements. However, the fact that the largest peaks are more pronounced in case of steering wheel angle cross-correlation results can be explained by the faster dynamics of the steering wheel angle. This was demonstrated in both studies using the amplitude spectra of lane position and steering wheel angle (see Figure 3.25 and Figure 3.39). Therefore, we can conclude that steering wheel angle cross-correlation functions are more sensitive to secondary task engagements compared to lane position cross-correlation functions. This can be seen clearly if we look at per-glance cross-correlation functions for the simplest conditions in both studies (AR and B): significant

peaks are detected for AR (navigation study) and B (iPod study) conditions in case of R_{sw} , but not in case of R_{lp} .

One question that can be asked here is as follows: why are there differences in the magnitudes of the observed cross-correlation peaks in the two studies? There are two main contributors to this result: interaction modality and driving environment. Both of these factors directly influence visual attention and driving performance while engaging in secondary tasks. However, we have to keep in mind that it is likely that these factors are coupled and that they cannot be considered entirely separately. It was shown in the previous studies that manual-visual interactions typically influence driving performance more strongly than predominantly visual or auditory interactions. Since driving performance directly contributes to the cross-correlation results we can expect that the observed differences between the studies would partially stem from the differences in interaction modality. Additionally, cross-correlation results depend on visual attention as well, thus any differences here would also affect the observed result. The other reason is that driving behavior depends largely on the road type and driving conditions. Obviously, driving on a busy city road creates a very different experience than driving on a highway during off-peak hours. For example, if we look at cumulative steering wheel angle cross-correlation functions for SV PND (navigation study) and D condition (iPod study) we can see that the magnitudes of the most prominent peaks are 10.65 degrees/second and 24.7 degrees/second, respectively. This result can be explained by the much larger average number of glances per segment for D (10.78) compared to SV (1.6). On the other hand, a higher per-glance cross-correlation peak was observed for SV (7.022 degrees/second) compared to D (2.136 degrees/second) condition. Since in this case we are observing the

individual instances of secondary task engagements (glances), one explanation is that this difference resulted from the overall difference in driving behavior between the two studies (despite the fact that on average D condition had longer glances (0.98 seconds) compared to SV (0.53 seconds)). In other words, each environment introduces some “baseline” variability in driving performance. This can be seen clearly by comparing the simplest conditions in these studies, namely, AR and B, respectively. We have to note here that the two conditions are not exactly the same. On the one hand, B condition represented true unencumbered driving, since no side task was involved. On the other hand, AR condition involved following navigation directions presented on the HUD, which captured at least some of drivers’ attention. However, the comparison is useful for indicating trends. Even though both conditions had very similar visual attention to the forward road ($PDT_{AR} = 96.48\%$, $PDT_B = 94.62\%$), they had very different impacts on driving. For example, variances of steering wheel angle were 4.81 degrees^2 and 0.23 degrees^2 for AR and B condition, respectively. We argue that the observed differences in driving were largely caused by the increased environment complexity that was present in the navigation study: two-lane streets, high traffic density, pedestrians, parked vehicles and short, narrow street segments with many consecutive turns. All these variables resulted in the higher expanded effort to maintain the vehicle in the center of the lane. Based on this we can assert that the driving environment is of considerable importance and undoubtedly has an influence on the cross-correlation results.

Effects of the Driving Environment

Based on the arguments provided in the previous section, we can state a new hypothesis (H4), which is concerned with the effects of driving environment. Namely, we

hypothesize that driving performance and cognitive load for the same secondary task would change between different driving conditions. Specifically, we expect that a more challenging driving environment may introduce larger effects on driving and cognitive load, which may be reflected in average-based and cross-correlation measures.

We propose to test the effect of the environment by comparing the results of two driving simulator studies which incorporate the same secondary task, but performed under different driving conditions. One study will be the iPod interaction study presented in this chapter (“Highway Driving and iPod Interactions”). The other study will include the same type of iPod interactions, except that they will be conducted in the city environment. Specifically, in this second study the participants will interact with an iPod while following a lead vehicle on a busy, straight city road. We will refer to this study as “City Driving and iPod Interactions” and it will be presented in Chapter 4. The fact that the two studies will have the same manual-visual task should enable us to precisely quantify the effect of the driving environment on both average-based driving performance measures and cross-correlation measures. Namely, given the equality of the secondary task engagements, it is expected that the amount of visual attention required to complete the tasks should be approximately the same between the two studies. Of course, it is possible that drivers may decide to protect the driving task by looking less at the iPod (given the increased complexity of the city environment). However, if the visual attention proves to be very similar between the two studies, any potential changes in the observed results can be attributed predominantly to the change in the driving environment. We expect to see larger variability in driving performance measures (specifically, steering

wheel angle and lane position) and larger cross-correlation results under city compared to the highway environment.

Exploring Underlying Mechanisms

Besides the effects of the driving environment, the next chapter will also explore influential variables (predictors) which can be used for explaining the underlying mechanisms that contribute to the observed cumulative and instance-based cross-correlation results (as proposed in hypothesis H3). In that respect, the two iPod studies (highway and city driving) lend themselves well, since they are fairly well controlled without any extraneous variables to account for (such as unexpected events and consecutive turns as in the navigation study). Therefore, they should allow easier identification of the most important predictors. We will refer to these two iPod studies as “reference” studies.

Testing Construct Validity with Physiological Measures

Even though subjective estimates are very informative and provide direct information about participants’ experiences, the problem is that they are not very objective. As we saw in the introduction, it was demonstrated in the literature that physiological measures can also be an effective way of characterizing changes in cognitive load. They are fairly difficult to be willingly impacted, thus providing a high level of objectiveness. It is for this reason that the study presented in the next chapter compares the cumulative cross-correlation results with two commonly used physiological measures: average heart rate and skin conductance. We expect that, similar to subjective estimates, a positive relationship will be revealed between the two. This will provide another source of support for construct validity of our cross-correlation method.

CHAPTER 4

MECHANISMS UNDERLYING CROSS-CORRELATION

RESULTS

The previous chapter provided a detailed description of the cross-correlation method as well as the results it produces based on two driving simulator studies. The current chapter will accomplish the following:

1. Present yet another driving simulator study which will demonstrate the effectiveness of the cross-correlation method in detecting changes in driving performance and cognitive load. Specifically, this study will be used for testing hypotheses H1 (quantification of cumulative effects of secondary task engagement), H2 (quantification of instance-based effects of secondary task engagement) and H_{RP} (ranking of the above cumulative and instance-based results).
2. Test the effect of driving environment on average-based and cross-correlation measures using the “reference” studies approach. Specifically, this chapter will compare a study which explores interactions with an iPod while driving in the city

environment with the study which was described in the previous chapter and explored the same type of interactions but in highway driving. If the effect of the environment is confirmed, it will provide support for hypothesis H4.

3. Use the results obtained from the “reference” studies to reveal the underlying variables (predictors) which have an important influence on the observed cross-correlation results. This will be used for testing hypothesis H3.

4.1 City Driving and iPod Interactions

Method

This study is very similar to the previous study (“Highway Driving and iPod Interactions”) in the sense that both involve the same type of secondary task: interactions with an MP3 player, specifically, an iPod Nano device. The experimental setup was exactly the same as in the previous study: the iPod was attached to a board paced on the right side of the steering wheel. This location allowed for easy manual-visual interaction without the need for large changes in gaze direction. A total of 12 participants (average age 19.6) participated in the study.

The primary task consisted of following a yellow lead vehicle which travelled at a constant speed of 40 MPH (64.4 km/h) (Figure 4.1). The simulated environment consisted of a straight city road with one lane in each direction, each 3.2 meters wide. We decided not to include any intersections, so as to assure uniform driving difficulty throughout the whole experiment. Both sides of the road were randomly populated with parallel-parked vehicles. The road was presented in daylight with frequent random traffic appearing both in the opposite lane (about 2 vehicles per second) and behind the

participant's vehicle. However, the ambient traffic did not interfere with either the lead vehicle or the participants' vehicle. The participants were instructed to follow the lead vehicle at a comfortable distance and to drive normally as they would in real life.



Figure 4.1 Simulated city environment.

The secondary task was exactly the same as in the previous study and it involved three levels of difficulty, which we will reiterate here briefly for completeness:

1. *No secondary task – baseline (B)*. This condition did not involve any interactions with the iPod – just following the lead vehicle.
2. *Easy iPod interaction (E)*. The participants were instructed to complete 10 simple actions with the iPod, such as playing the current song, rewinding a song, and increasing/decreasing volume.
3. *Difficult iPod interaction (D)*. The participants were instructed to find and play 10 songs from a list of 347 songs. Both the list and the sought songs were sorted

alphabetically. This simplified the task, since it required scrolling in only one direction.

In both easy and difficult conditions, the participants were instructed which action to perform using a computer voice. The participants had 40 seconds to complete each task.

As in the previous study, we chose a within-subjects factorial design experiment with the interaction type as the primary independent variable, *Int*. The levels of *Int* were B, E and D and their order was counterbalanced among the participants in order to circumvent the learning effect. The following dependent variables were collected: PDT on the forward road, average glance duration, average number of glances, average-based driving performance measures expressed through variances of steering wheel angle, lane position and velocity, average velocity and subjective estimates of cognitive load (using the NASA-TLX questionnaire). As we discussed in Chapters 1 and 2, other researchers have indicated the usability of physiological measures for detecting changes in cognitive load. It is for this reason that we decided to include physiological measures in this study as well, besides the variables listed above. Specifically, we collected average heart rate and skin conductance (see Appendix B for a description of our physiological measurements monitor).

Since there were 10 interactions with the iPod, we divided our experiment in ten 40-second-long segments. This segmentation was performed for the B condition as well, so we would be able to make direct comparisons with the other conditions. All of our dependent variables were calculated for each experimental segment.

General Results

Visual attention analyses will be presented first. A repeated-measures one-way ANOVA indicated a significant main effect of the interaction type on PDT on the forward road ($F(2,22)=115.279$, $p<0.0001$) (Figure 4.2). Post-hoc comparisons indicated highly significant differences between all possible pairs ($p<0.0001$).

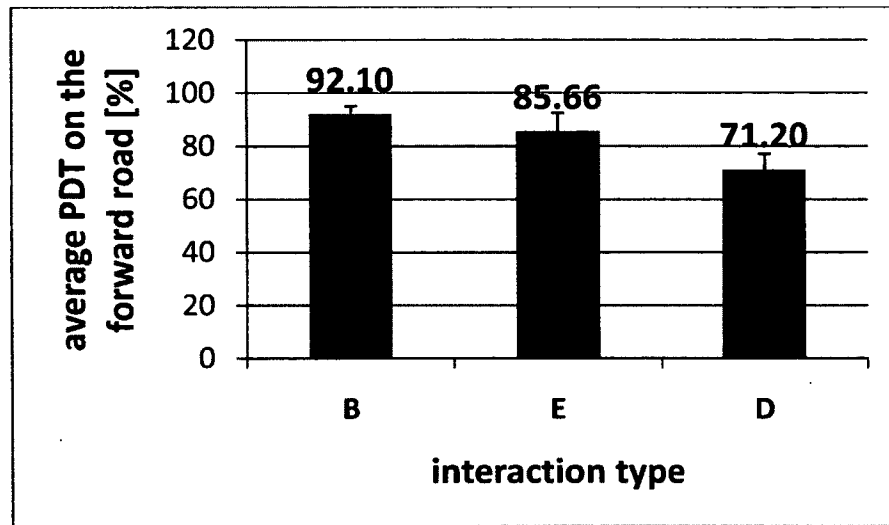


Figure 4.2 Average PDT on the forward road.

As we can see in Figure 4.2, the participants spent 92.1%, 85.7% and 71.2% of time looking at the forward road for B, E and D condition, respectively. If we use these percentages to calculate the amount of time participants spent looking away from the road for each minute of driving, it would amount to 17.28 seconds for D, 8.6 seconds for E, and only 4.74 seconds for B task. These numbers indicate that as the complexity of the secondary task increased, visual attention shifted away from the road more.

More details about changes in visual attention can be obtained if we look at average duration and number of glances directed away from the road for each interaction type. The same set of statistical analyses as in the previous study was performed here as well. Figure 4.3 left shows the average glance durations to be 0.63, 0.65 and 0.87 seconds

for B, E, and D condition, respectively. A one-way ANOVA indicated a significant main effect of the interaction type on glance duration ($F(2,3072)=99.9402$, $p<0.0001$). Post-hoc pairwise comparisons revealed significant differences between D and B ($p<0.0001$) and D and E ($p<0.0001$), but not between E and B ($p=0.6108$) conditions. The same conclusion was obtained using a non-parametric Kruskal-Wallis test: significant main effect ($\chi^2=106.1096$, $p<0.0001$) and significant differences between all pairs ($p<0.0001$) except E and B ($p=0.9657$).

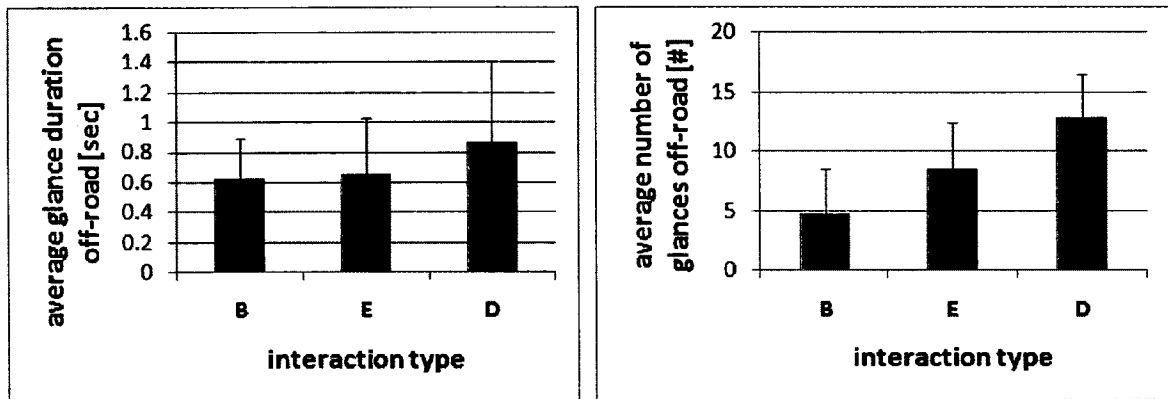


Figure 4.3 Average duration (left) and number (right) of glances directed off-road.

As we can see in the right graph of Figure 4.3, the average number of glances directed off-road is 4.75, 8.5 and 12.77 for B, E and D, respectively. A one-way ANOVA indicated a significant main effect of the interaction type on number of glances ($F(2,352)=131.5559$, $p<0.0001$). Post-hoc comparisons indicated differences between all pairs ($p<0.0001$). A non-parametric Kruskal-Wallis test also indicated a significant main effect ($\chi^2=156.5675$, $p<0.0001$) and significant pairwise differences between all pairs ($p<0.0001$) of tasks.

Next, we analyzed subjective estimates of cognitive load obtained using the NASA-TLX questionnaire (Figure 4.4). A significant main effect of interaction type was

detected using a repeated-measures ANOVA ($F(2,22)=32.072$, $p<0.0001$). Pairwise comparisons indicated differences between all pairs: B and E ($p<0.0001$), B and D ($p<0.0001$) and E and D ($p=0.002$). We can say that the subjective estimates agree with the visual attention results: participants judged D condition to be the most difficult, followed by E and B conditions.

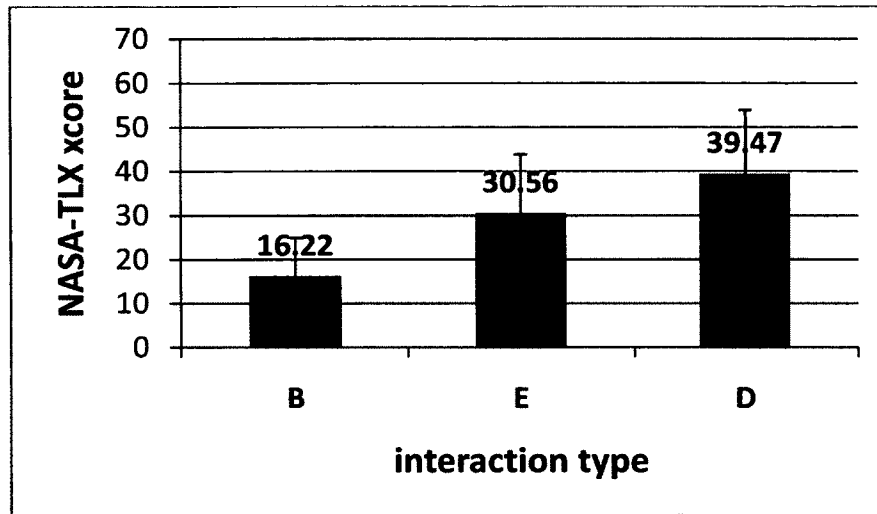


Figure 4.4 Average NASA-TLX score.

The results obtained through the subjective estimates of cognitive load were also confirmed by one physiological measure: average skin conductance. Figure 4.5 shows the average values for both skin conductance and heart rate.

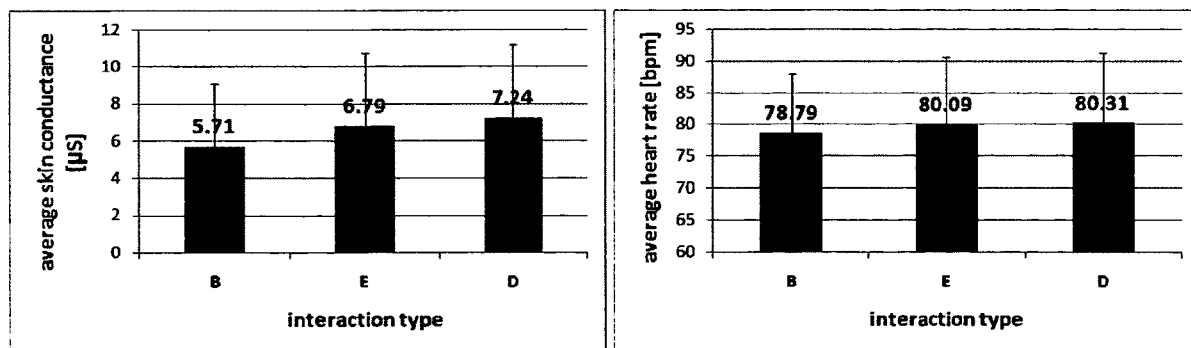


Figure 4.5 Average physiological measures: skin conductance (left) and heart rate (right).

One of the participants accidentally disconnected the heart rate electrode while driving the simulator, so the heart rate data was not available in that case. Thus, heart rate is based on 11, while skin conductance is based on 12 participants.

A repeated-measures ANOVA indicated a significant main effect of interaction type on skin conductance ($F(2,22)=6.451$, $p=0.006$). Post-hoc pairwise comparisons revealed a highly significant difference between B and D conditions ($p=0.003$) and a marginally significant difference between B and E conditions ($p=0.053$). No difference has been observed between E and D conditions ($p=0.299$). If we look at the skin conductance values, we can see that participants experienced the lowest workload during the B condition, followed by E and D conditions. Even though the difference between E and D is not statistically significant, we can clearly see that skin conductance indicates the same trend observed with subjective estimates of cognitive load. No significant effect of interaction type has been observed on average heart rate ($p>0.05$).

Finally, we analyzed the effects of the three interaction types on driving performance using average-based measures. Figure 4.6 shows average variances of lane position (upper left), steering wheel angle (upper right), velocity (lower left) and average velocity (lower right). For each measure we performed a repeated-measures ANOVA with interaction type as the independent variable. Conditional on the significant main effect, we also performed pairwise comparisons. Table 4.1 outlines these results. As we can see a significant main effect of interaction type has been observed for all variables except lane position variance. Furthermore, in case of variances of steering wheel angle and velocity, post-hoc comparisons revealed significant differences between all possible

pairs ($p < 0.05$). In case of average velocity, the only significant difference has been observed between E and D conditions.

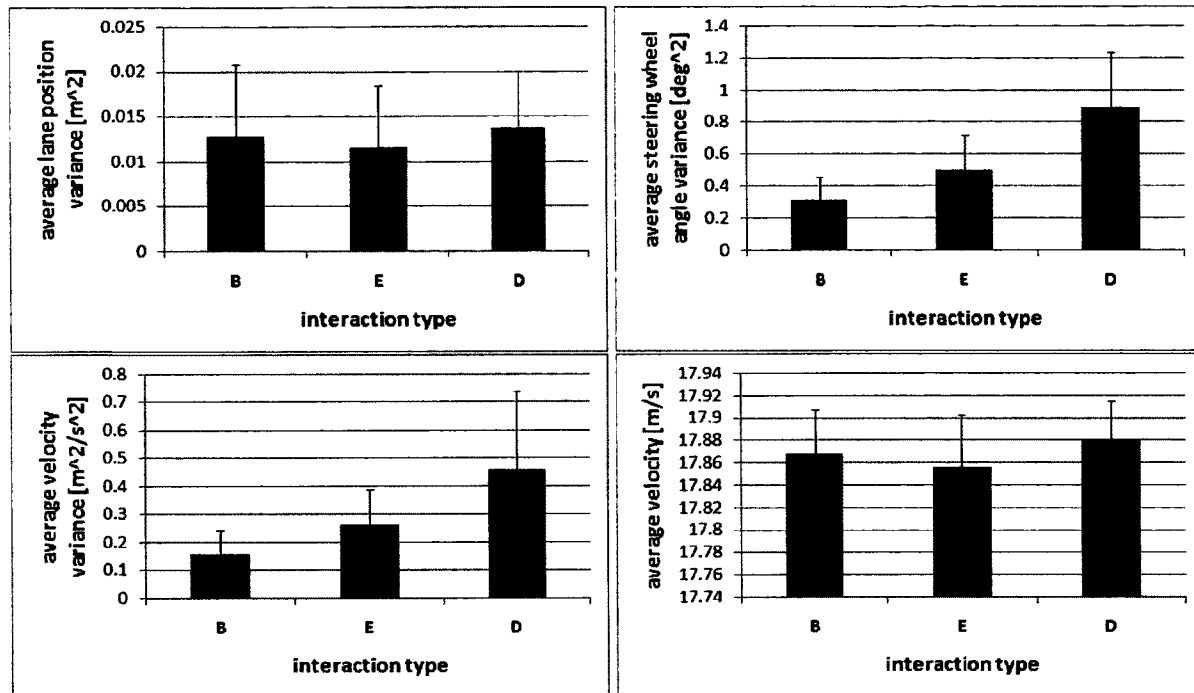


Figure 4.6 Average variances of lane position (upper left), steering wheel angle (upper right), velocity (lower left) and average velocity (lower right).

Dependent variable	F-value	p-value	p-values for pairwise comparisons		
			B - E	B - D	E - D
Lane position variance	$F(2,22) = 1.922$	0.170	N/A	N/A	N/A
Steering wheel angle variance	$F(2,22) = 27.401$	<0.0001	0.007	<0.0001	<0.0001
Velocity variance	$F(2,22) = 10.253$	0.001	0.01	0.002	0.036
Average velocity	$F(2,22) = 3.552$	0.046	0.236	0.212	0.016

Table 4.1 Statistical analyses of average-based driving performance measures.

Based on these results we can conclude that iPod interactions in a busy city environment resulted in higher variability of the steering wheel angle compared to just

driving (B), which can be explained by the participants exerting higher effort to keep the vehicle in the center of the lane. We have to remind ourselves that the road consisted of two 3.2 meters wide lanes with high volume of ambient traffic and parked vehicles on both sides of the road. It is likely that this demanding driving environment gave participants more incentive to work harder in order to avoid collisions with the surrounding objects. This may also explain the lack of significant main effect for lane position. Nevertheless, it does not mean that the effect of interactions is not present, merely that it was not detected using the average-based approach. Increased velocity variance is another indicator that the participants had harder time keeping their speed constant as the difficulty of the secondary task increased.

Cross-Correlation Results

Cumulative steering wheel angle and lane position cross-correlation results are presented in Figures 4.7 and 4.8, respectively. In both figures solid lines represent cross-correlation functions, while dotted lines represent their significance levels of 0.05.

Figure 4.7 shows that significant peaks exist in $R_{sw}[lag]$ for all interaction types. The peaks represent the average absolute cumulative angular change (AVC) on the steering wheel while performing each interaction task with the iPod. They indicate that on average there is a larger cumulative change in the steering wheel angle following glances directed away from the forward road than in usual circumstances. The significant peak is also present in the B condition, which is the result of occasional glances towards the speedometer, steering wheel or dashboard. However, its magnitude is considerably smaller compared to D and E conditions. The most prominent peaks occur on average at the lags of 0.5 seconds for D, 0.4 seconds for E and 0.6 seconds for B condition.

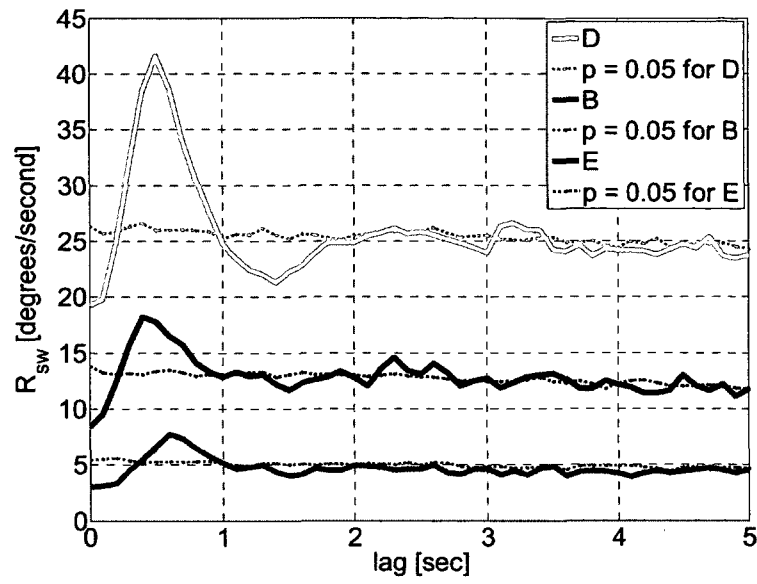


Figure 4.7 Cumulative steering wheel angle cross-correlation functions calculated for B, E and D conditions.

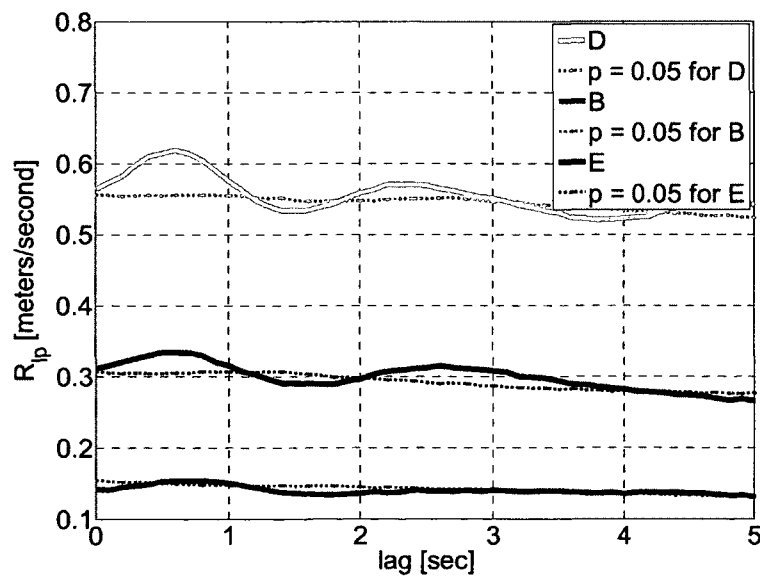


Figure 4.8 Cumulative lane position cross-correlation functions calculated for B, E and D conditions.

Figure 4.8 shows that significant peaks also exist in case of cumulative lane position cross-correlation results for all three interaction types. The most prominent peaks appear at 0.6, 0.5 and 0.6 seconds for D, E and B condition, respectively.

We can see that, similar to the previous study, in both cumulative steering wheel angle and lane position cross-correlation functions some distant cross-correlation peaks occur on average more than 2 seconds away from the edge of the glance. If we take into account that the speed limit in this study was 40 MPH, we can calculate that the car would travel 35.76 meters during the interval of 2 seconds. Given that the driving environment was populated both with ambient traffic and parked vehicles on both sides of the road, it is likely that the necessary correction of the car's position in the lane would have to be applied much earlier in order to avoid a collision. The most prominent peaks also support this assertion, since their magnitudes are larger compared to the magnitudes of the distant peaks.

As we had a chance to see so far, significant cumulative effects of looking away from the road were detected for all interaction types and for both steering wheel angle and lane position. This is an important result. However, we would also like to know whether those effects are different between the three interaction types and how they rank. To accomplish this, as with the previous study, we used the two comparison procedures presented in Section 3.1.5. The results are outlined in Table 4.2.

Using both approaches we detected a significant main effect ($p < 0.0001$) of the interaction type on the cumulative cross-correlation results for both steering wheel angle and lane position. Furthermore, post-hoc pairwise comparisons revealed differences between all possible pairs ($p < 0.0001$). Given that all differences are statistically

significant, we can conclude that the largest cumulative impact on driving over the course of interaction with the iPod was introduced by D condition, followed by E and B conditions. To obtain a sense of how large the effect is, we can compare the magnitudes of the most prominent peaks between individual conditions. Since B condition represented true unencumbered driving, it makes an ideal reference for comparisons. If we take R_{sw} as an example, we can see that D produced $41.6/7.714 = 5.36$ times larger cumulative effect than B condition. Similarly, if we compare E and B conditions, we can see $18.19/7.714 = 2.36$ times larger effect in case of E condition. Finally, when comparing D and E conditions alone, we can see that D produced $41.6/18.19 = 2.29$ times larger effect. The effect sizes can be calculated analogously for R_{lp} results.

Cumulative steering wheel angle cross-correlation	Comparing Highest Peaks			Comparing Areas Below Curves		
	Main effect of <i>Int</i>		p < 0.001	Main effect of <i>Int</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
Cumulative lane position cross-correlation	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001
	Main effect of <i>Int</i>		p < 0.001	Main effect of <i>Int</i>		p < 0.001
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001	p < 0.001

Table 4.2 Results of statistical comparisons between cumulative cross-correlation functions for B, E and D conditions.

In order to analyze how the conclusions obtained from the cumulative cross-correlation results compare to other estimates of cognitive load, we turn to subjective and physiological measures. As we had a chance to see in the introduction both of these

measures describe overall changes in cognitive load. Therefore, their comparison with our cumulative cross-correlation results is sound. Figure 4.9 demonstrates positive relationships between the magnitudes of the most prominent cumulative steering wheel angle and lane position cross-correlation peaks versus NASA-TLX results. Using simple linear fitting, we obtained very strong positive relationships in both cases (coefficients of determination are $R^2 \geq 0.88$), which indicate that all of these measures lead to the same conclusions with respect to the overall cognitive load changes: D produces the highest impact, followed by E and B conditions.

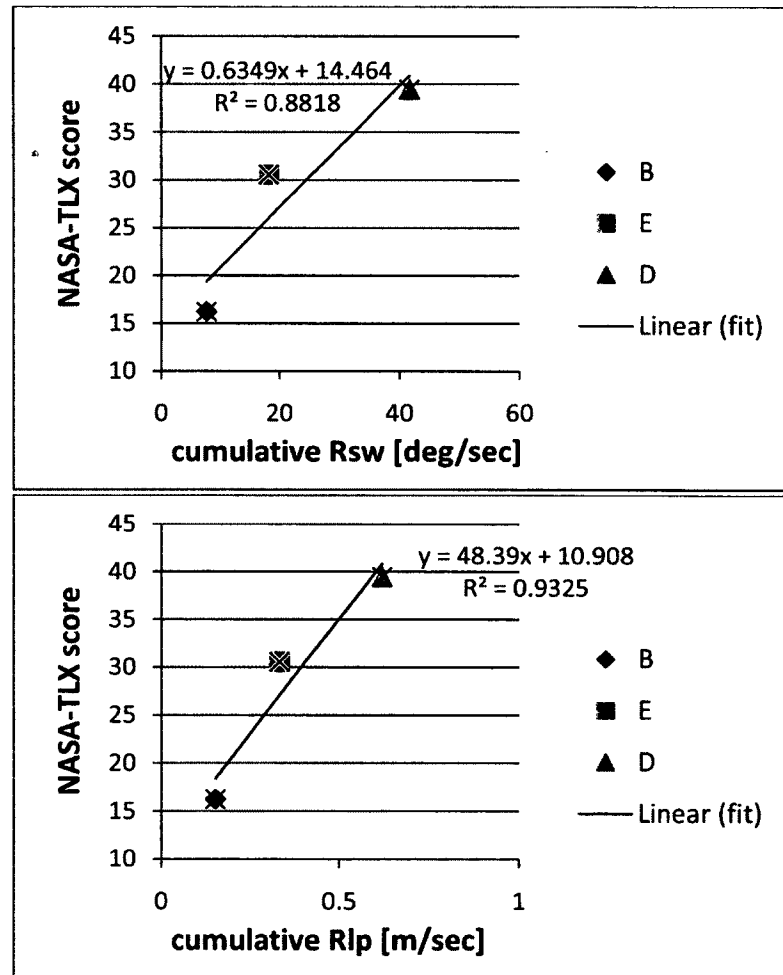


Figure 4.9 Magnitudes of the most prominent peaks of cumulative cross-correlation

functions R_{sw} and R_{lp} vs. NASA-TLX score for B, E and D conditions.

Since in this study we detected a significant main effect of the interaction type with the iPod on average skin conductance, we compared those results to the ones obtained using our cumulative cross-correlation measure. Figure 4.10 shows strong positive relationships ($R^2 \geq 0.81$) between the magnitudes of the most prominent cumulative cross-correlation peaks for both steering wheel angle and lane position versus average skin conductance.

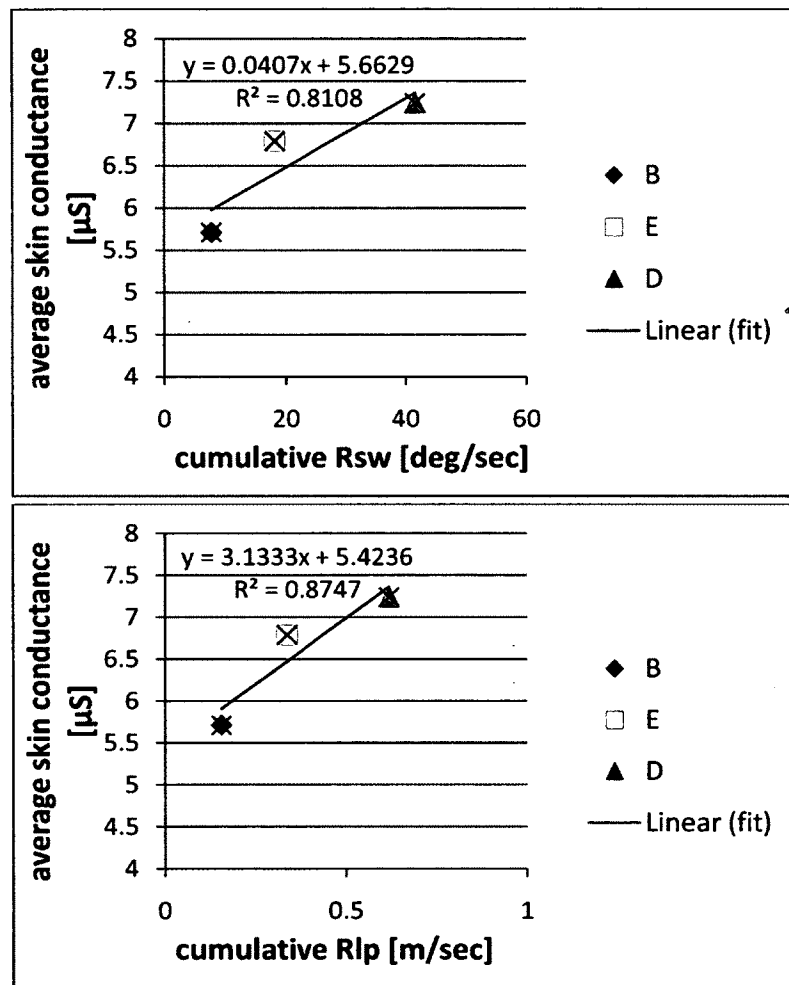


Figure 4.10 Magnitudes of the most prominent peaks of cumulative cross-correlation functions R_{sw} and R_{lp} vs. average skin conductance for B, E and D conditions.

We have to recall here that the difference between D and E conditions in case of average skin conductance was not significant ($p=0.299$, see Figure 4.5). This indicates

that average skin conductance did not provide high enough sensitivity to detect this difference, while our method did. Nevertheless, the comparisons presented in Figure 4.10 are still valuable, since they indicate important trends which lead to the same overall conclusions between the two measures.

Now that we understand the cumulative effects of iPod interactions on cognitive load, we can also perform a more fine-grained analysis by observing the impacts of individual instances of interactions (off-road glances in our case). Figures 4.11 and 4.12 depict per-glance steering wheel angle and lane position cross-correlation results for all interaction types, respectively. As before, solid lines represent cross-correlation functions, while dotted lines represent their significance levels of 0.05.

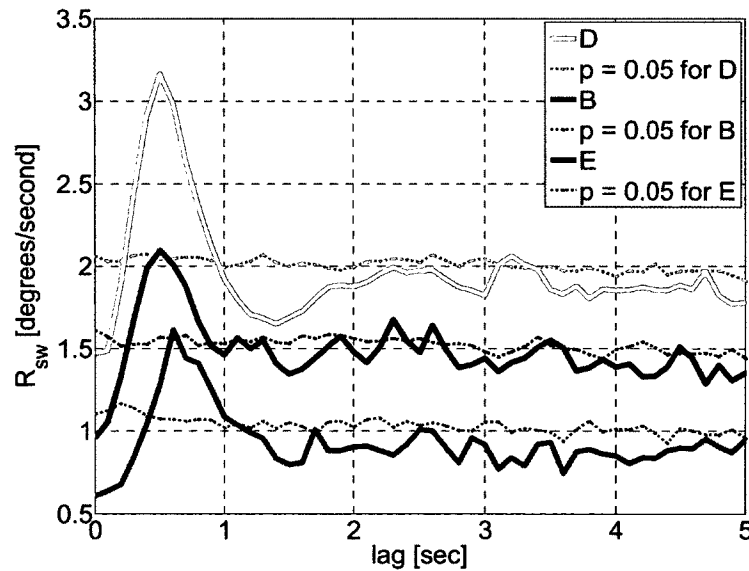


Figure 4.11 Per-glance steering wheel angle cross-correlation functions calculated for B, E and D conditions.

As we can see in Figure 4.11 significant effects of individual glances directed off-road exist for all three interaction types, which are judged by the highly significant

peaks in the per-glance steering wheel angle cross-correlation functions. These peaks indicate the average absolute amount of angular change (AVC) on the steering wheel resulting from an average glance directed off-road. The lags of the most prominent peaks are 0.5 seconds for D and E, and 0.6 seconds for B condition. It is very interesting to see that the impact of an occasional glance directed off-road exists in the B condition as well.

Similar results are obtained in case of lane position. Namely, Figure 4.12 demonstrates significant per-glance lane position cross-correlation peaks for all three conditions. These peaks indicate the average amount of change in the lane position contributed by an average glance directed away from the road. The most prominent peaks occur at the lags of 0.6 seconds for D, 0.7 seconds for E and 0.8 seconds for B condition.

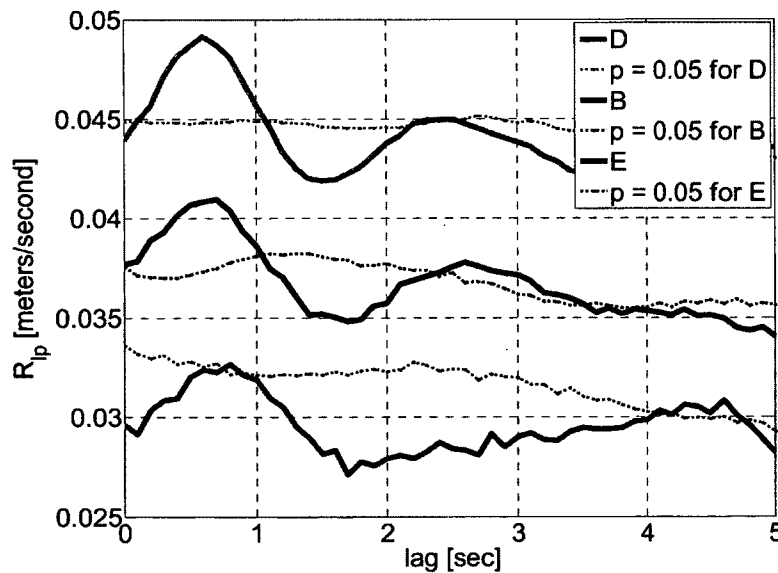


Figure 4.12 Per-glance lane position cross-correlation functions calculated for B, E and D conditions.

Finally, we performed statistical comparisons to determine whether there exist any differences in the effects of average glances on the per-glance cross-correlation

results between the three conditions. Table 4.3 outlines the obtained results. As we can see, both procedures (highest peaks and areas below the curves) indicated significant main effects ($p < 0.0001$) of interaction type for both per-glance steering wheel angle and lane position cross-correlation functions. All pairwise comparisons indicated significant differences as well ($p < 0.0001$). These results demonstrate that the impacts on driving resulting from individual glances directed away from the road are different and depend on the difficulty of the interaction, with D condition producing the largest impact, followed by E and B conditions. This ranking agrees with the one obtained with the cumulative cross-correlation functions.

Per-glance steering wheel angle cross-correlation	Comparing Highest Peaks			Comparing Areas Below Curves		
	Main effect of <i>Int</i>		$p < 0.001$	Main effect of <i>Int</i>		$p < 0.001$
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
Per-glance lane position cross-correlation	$p < 0.001$	$p = 0.0039$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Main effect of <i>Int</i>		$p < 0.001$	Main effect of <i>Int</i>		$p < 0.001$
	Pairwise comparisons			Pairwise comparisons		
	B-D	B-E	D-E	B-D	B-E	D-E
Per-glance steering wheel angle cross-correlation	$p < 0.001$	$p = 0.0018$	$p < 0.001$	$p < 0.001$	$p = 0.0006$	$p < 0.001$

Table 4.3 Results of statistical comparisons between per-glance cross-correlation functions for B, E and D conditions.

By comparing the magnitudes of the most prominent peaks we can obtain the relative size of the effect contributed by an average glance. For example, if we use per-glance steering wheel angle cross-correlation functions we can see the following effects: D produced $3.165/1.613 = 1.96$ times larger impact compared to B and $3.165/2.096 =$

1.51 times larger impact compared to E condition. If we compare E and B conditions, we can see $2.096/1.613 = 1.3$ times larger impact in case of E condition.

The fact that we revealed significant differences between interaction types on the per-glance basis is a very important one. Besides knowing that the three interaction types are different regarding their cumulative effects on driving and cognitive load, this indicates that the differences exist at a much lower level as well, namely, at the level of an average glance. This is an important finding, because it provides a new insight into the performed activity, in this case interactions with the iPod. In other words, we observed that the D condition resulted in the largest cumulative effect, which could have been expected given the associated level of involvement. However, it was not obvious that the D condition also produced the largest effects in the individual instances of interaction.

General Conclusions

As we had a chance to see in the previous section, our cross-correlation method successfully detected both cumulative and instance-based (per-glance in our case) influences on cognitive load. In each case we detected significant impacts of looking away from the road resulting from iPod interactions, which was indicated by statistically significant cross-correlation peaks. These are important results because they demonstrate when the influences occur (lag) as well as how large they are (magnitude of a significant peak). This is possible because our method takes time into account. Conversely, average-based measures analyze an experimental condition as a whole, thus characterizing it with only a single value. Furthermore, we demonstrated significant differences between interaction types based on the steering wheel angle and lane position cross-correlation results (both cumulative and per-glance), which allowed us to rank the effects of those

interactions produced on driving and cognitive load. The significant ranking that we obtained also allowed us to calculate the relative sizes of the effects between the three interaction types.

It is worth noting that the average variance of lane position did not even detect the main effect of interaction type (see Table 4.1). This indicates the complete lack of sensitivity that the average variance of lane position demonstrated in this study. On the other hand, our method demonstrated higher sensitivity by detecting both the main effect of the interaction type as well as all pairwise differences.

Based on all of the above results we can conclude that hypotheses H1, H2 and H_{RP} are supported. Furthermore, we demonstrated that our cross-correlation method is capable of providing more sensitivity to changes in cognitive load compared to average-based driving performance measures.

This study provides ample evidence which supports construct validity of our method:

1. A significant main effect of the interaction type was detected in case of the following average-based driving performance measures: average variances of steering wheel angle and velocity, and average velocity. Post-hoc pairwise comparisons indicated significant differences between all interaction types for the first two variables. The ranking based on those differences matches the ranking obtained using our cross-correlation method.
2. A significant main effect of the interaction type was detected for the subjective estimates of cognitive load based on the NASA-TLX questionnaire. Significant

differences were detected between all interaction types and the ranking matched the one obtained with our method. Additionally, we demonstrated through linear regression models a very strong positive relationship between the two types of measures (see Figure 4.9).

3. Finally, a significant main effect of the interaction type was detected for one physiological estimate of cognitive load, namely, average skin conductance. Even though this measure was not sensitive enough to detect the difference between D and E conditions, the magnitudes of the most prominent cumulative cross-correlation peaks and the average values of skin conductance for different interaction types showed a strong positive relationship (see Figure 4.10).

4.2 Observing Effects of Driving Environment through

Reference Studies

This section investigates the effect of the driving environment and how it was reflected in visual attention, average-based driving performance measures and cross-correlation results. As we hypothesized (H4) in Section 3.2.3, we expect that driving environment may have a significant effect on all of the above results. However, based on the results obtained in the previous studies, we expect that our method may again provide more sensitivity compared to average-based measures. We tested hypothesis H4 by comparing the results of the two reference studies: “Highway Driving and iPod Interactions” and “City Driving and iPod Interactions.”

Both reference studies incorporated exactly the same secondary task: easy and difficult interactions with the iPod while driving. In both cases we also introduced a

baseline condition, which did not include any interactions and thus represented true unencumbered driving. The only aspect that changed between the two studies was the driving environment. In the first study the participants drove on a wide, three-lane highway road with light ambient traffic. Conversely, in the second study the participants drove on a narrow, two-lane city road with high volume of ambient vehicles as well as parallel-parked vehicles on both sides of the road. As we can see, the change in the environment was significant. By conducting these two reference experiments we have the opportunity to observe and explain the changes introduced by the environment in both cross-correlation and average-based measures. Since both reference studies were well controlled (we made an effort to minimize the number of confounding variables) and the driving environment was the only difference between the two, we can be fairly confident that it affected the majority of the differences in the observed results.

4.2.1 Effects of Driving Environment on Visual Attention

In this section we will observe how the change in driving environment between two reference studies influenced visual attention.

If we take a look at the average PDT directed to the forward road, we can see that it is practically the same between the two studies. Figure 4.13 illustrates this. Dark gray indicates city driving, while light gray indicates driving in the highway environment. We conducted a two-way ANOVA in order to test whether significant differences exist between the two studies regarding PDT to the forward road. We used PDT as our independent variable, while *interface type* (levels: B, E, D) and *environment type* (levels: highway, city) served as independent variables. We also included an interaction term

interface \times *environment* in our model in order to check for the potential interaction between the two. Please note that for the purposes of statistical analyses we will refer to the three interaction types with the iPod as “*interface*” in order to distinguish it from the statistical interaction that may exist between the two independent variables: *interface* and *environment*. The results indicated a significant main effect of *interface type* ($F(2,66)=75.8384$, $p<0.0001$). No significant main effect has been observed for *environment type* ($F(1,66)=0.775$, $p=0.3819$). Finally, no significant interaction between *interface* and *environment* has been detected ($F(2,66)=0.4156$, $p=0.6617$). Based on these results we can conclude that the participants allocated approximately the same amount of visual attention towards the secondary task in each driving environment.

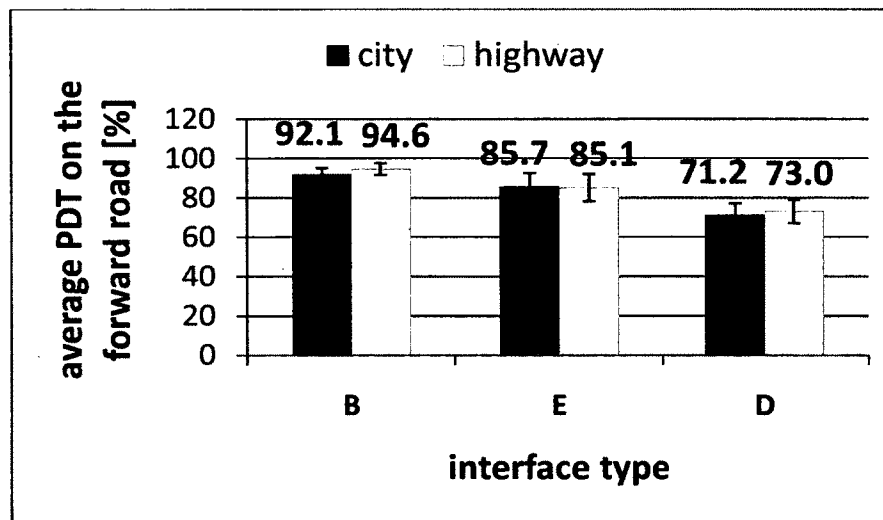


Figure 4.13 Average PDT on the forward road observed in city and highway driving while interacting with the iPod.

Since this particular iPod variant cannot be operated without looking at the device (it is necessary to observe the contents of the LCD screen and the buttons do not provide a tactile feedback when operated), visual attention directed to the road represents a very good proxy for how the participants actually interacted with the device. This

information is certainly valuable. However, in order to obtain a low level insight into these interactions (operating different interface types), we have to look at fine grained descriptors, specifically average glance duration and number of glances.

Figure 4.14 shows the average glance duration calculated for each *interface* and *environment type*. Again, dark gray indicates city, while light gray indicates highway environment.

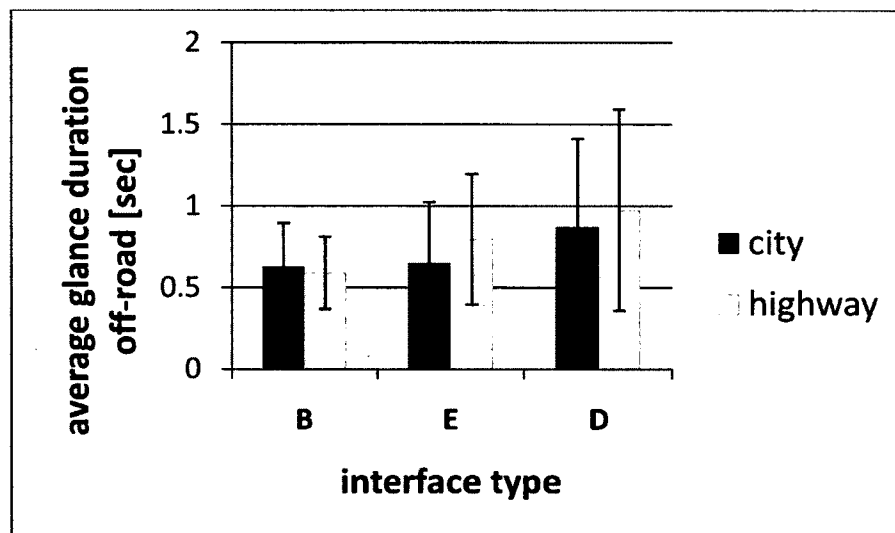


Figure 4.14 Average glance duration directed off-road in city and highway driving while interacting with the iPod.

We conducted a two-way ANOVA to explore the effects of the two environments on glance duration. Thus, glance duration was the dependent variable, while the independent variables were the same as with the PDT. The results for the ANOVA indicated a significant main effect for the *interface type* ($F(2,5605)=195.4455$, $p<0.0001$), a significant main effect for the *environment type* ($F(1,5605)=25.0659$, $p<0.0001$) and a significant interaction between the above independent variables ($F(2,5605)=12.2515$, $p<0.0001$). In order to determine the levels of *interface type* at

which the differences in *environment type* occur, we proceeded with pairwise comparisons. We obtained significant differences in glance duration between the two *environment types* for each *interface type*: B ($F(1,979)=5.9253$, $p=0.0151$), E ($F(1,1870)=64.9796$, $p<0.0001$) and D ($F(1,2756)=22.2041$, $p<0.0001$).

Figure 4.15 shows the average number of glances obtained for each *environment* and *interface type*.

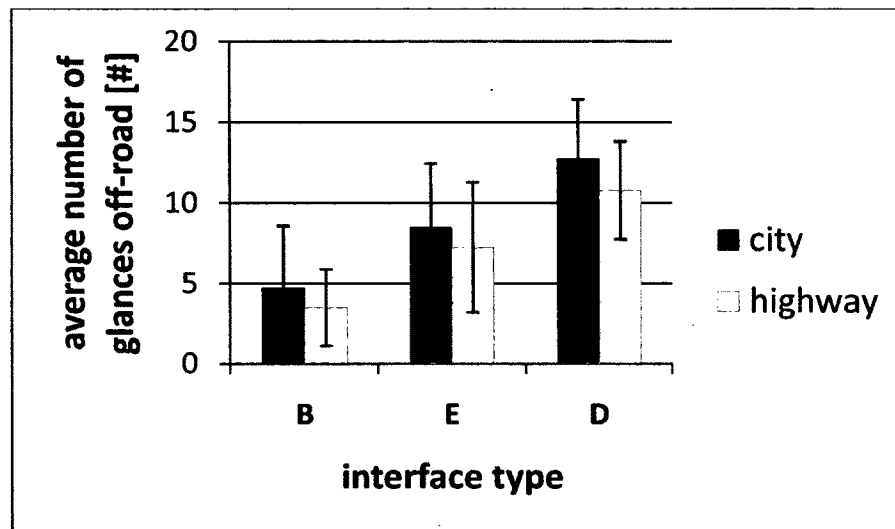


Figure 4.15 Average number of glances directed off-road in city and highway driving while interacting with the iPod.

As before, we conducted a two-way ANOVA with the number of glances as the dependent variable. The results revealed a significant main effect of the *interface type* ($F(2,704)=277.2689$, $p<0.0001$), a significant main effect of the *environment type* ($F(1,704)=31.9781$, $p<0.0001$) and a non-significant interaction between the two variables ($F(2,704)=0.8413$, $p=0.4316$). Pairwise comparisons between two *environment types* within each level of *interface type* showed significant differences in all cases: B

($F(1,236)=9.1796$, $p=0.0027$), E ($F(1,236)=5.92$, $p=0.0157$) and D ($F(1,232)=20.3944$, $p<0.0001$).

If we would look at PDT alone, we would conclude that driving environment did not produce any effect on visual attention. However, based on the results obtained from average glance duration and number of glances we can conclude that the environment did actually influence visual attention significantly. Namely, if we consider glance duration alone, we can see that during highway driving the participants made longer glances off-road compared to when they drove in the city. On the other hand, the participants glanced less frequently (smaller number of glances) away from the road in the case of highway road compared to city road. In other words, the participants cast larger number of shorter glances away from the road in the city environment and smaller number of longer glances in the highway environment. Based on these results we can conclude that the participants considered the highway environment to be more “forgiving” towards reduced visual attention (at least at the level of individual glances), as opposed to the city environment. These results also explain why the overall visual attention to the forward road appeared to be the same, as judged by PDT. Therefore, we can conclude that H4 is in fact satisfied with respect to visual attention results.

4.2.2 Effects of Driving Environment on Average-Based Driving Performance Measures

We can also look at the environmental impact through average-based driving performance measures, specifically variances of lane position, steering wheel angle and velocity. Note that comparing average velocity between the two driving environments is

not possible, given different speed limits (highway = 55 MPH, city = 40 MPH). Figures 4.16, 4.17 and 4.18 show the average variances of steering wheel angle, lane position and velocity for the two reference studies.

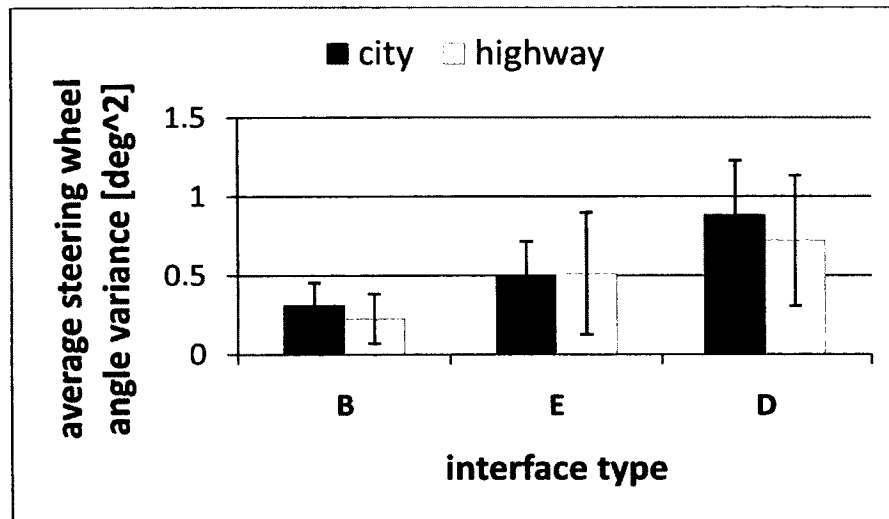


Figure 4.16 Average steering wheel angle variance in city and highway driving while interacting with the iPod.

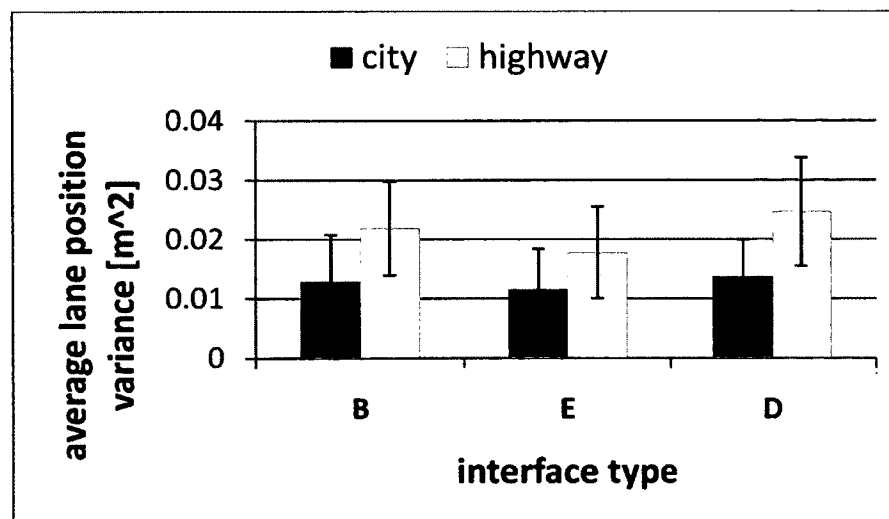


Figure 4.17 Average lane position variance in city and highway driving while interacting with the iPod.

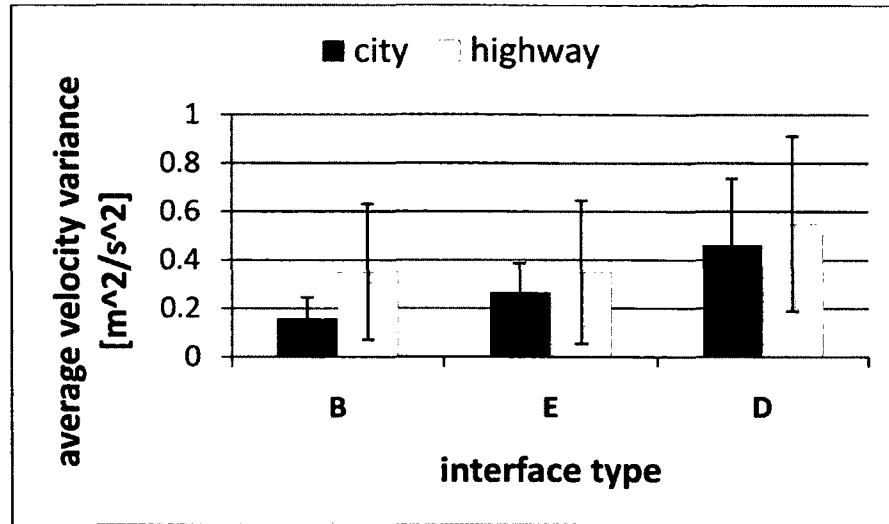


Figure 4.18 Average velocity variance in city and highway driving while interacting with the iPod.

Regarding steering wheel angle variance, a two-way ANOVA indicated a significant main effect of *interface type* ($F(2,66)=19.6844$, $p<0.0001$), a non-significant effect of *environment type* ($F(1,66)=1.3282$, $p=0.2533$) and a non-significant interaction between the two ($F(2,66)=0.5552$, $p=0.5766$). Regarding lane position variance, a two-way ANOVA indicated a significant main effect of *environment type* ($F(1,66)=23.2034$, $p<0.0001$), but not *interface type* ($F(2,66)=2.111$, $p=0.1292$) or the interaction between the two independent variables ($F(2,66)=0.5693$, $p=0.5687$). Finally, regarding velocity variance, a two-way ANOVA indicated a significant main effect of *environment type* ($F(1,66)=4.0704$, $p=0.0477$), a significant main effect of *interface type* ($F(2,66)=6.3766$, $p=0.0029$), and a non-significant interaction between the two ($F(2,66)=0.3434$, $p=0.7106$). As we can see, variances of lane position and velocity were more sensitive to changes in the driving environment compared to steering wheel angle.

Since average variances of both lane position and velocity detected a significant effect of *environment type*, we proceeded with pairwise comparisons within

interface types. Significantly larger lane position variances were detected while driving on the highway for all *interface types*: B ($F(1,22)=7.8181$, $p=0.0105$), E ($F(1,22)=4.3797$, $p=0.0481$) and D ($F(1,22)=11.6291$, $p=0.0025$). This agrees with Zhang et al. [30], who found that the variation of lane position was larger on highways than on rural roads. In case of velocity, a significantly larger variance has been observed in highway driving for B ($F(1,22)=5.1677$, $p=0.0331$), but not in case of E ($F(1,22)=0.8358$, $p=0.3705$) or D ($F(1,22)=0.4583$, $p=0.5055$) conditions.

Even though steering wheel angle variance did not detect a significant effect of the environment, it is interesting to note that it was typically higher in the city (at least for B and D, see Figure 4.16). On the other hand lane position variance was higher on the highway for all *interface types*. This suggests that the participants expended higher effort in order to keep the vehicle in the middle of the lane when driving in the city environment. This finding is sound given the high volume of ambient traffic, narrower streets and parked vehicles present in the city environment. Nevertheless, we do not possess a specific evidence for this argument given the lack of significant effect on steering wheel angle variance. We can also see that participants' velocity varied more while driving on the highway. It is possible that participants found the highway road less demanding and therefore they invested less effort to keep a constant distance (gap) behind the lead vehicle.

Based on these results we can say that hypothesis H4 is mostly supported by the average-based driving performance measures, specifically variances of lane position and velocity. However, the variance of steering wheel angle was not sensitive enough to detect differences between the two driving environments.

4.2.3 Effects of Driving Environment on Cross-Correlation

Results

The effects of the driving environment can be seen clearly in our cross-correlation results. Larger number of glances directed away from the road in the city environment produced larger cumulative effects on steering wheel angle cross-correlation results compared to the highway (compare Figures 3.33 and 4.7). Please note, however, that the number of glances is not the sole contributor to the cumulative result, since we demonstrated using our per-glance cross-correlation results that the effects of individual glances depend on the performed activity (in other words, they differ between B, E and D conditions). We can see the same trend regarding per-glance steering wheel angle cross-correlation functions: higher cross-correlation peaks in the city compared to the highway. However, the effect was opposite in case of cumulative and per-glance cross-correlation functions for lane position: higher peaks have been observed on the highway compared to the city. This result provides evidence for our argument stated in the previous section: the participants were expending higher effort on the steering wheel in order to keep the vehicle in the center of the lane on the city road (due to environment complexity) which resulted in smaller changes in lane position. On the other hand, on the highway, the participants invested less effort on the steering wheel (due to simpler driving environment) which resulted in larger changes in the lane position. This result is very important, because it provides another source of support that our method can provide more sensitivity compared to average-based driving performance measures.

In order to test the significance of the observed effect of driving environment on our cross-correlation results we conducted several two-way ANOVAs. We used the

magnitudes of the most prominent cross-correlation peaks (both cumulative and per-glance for steering wheel angle and lane position) as our dependent variable, while *environment type*, *interface type* and *interface* \times *environment* were used as independent variables. Table 4.4 gives the details of the statistical analyses. As we can see for each cross-correlation result there is a significant effect of all independent variables. It is valuable to note that the effect of the environment was always significant, which confirmed our expectations based on the obtained results.

Cross-correlation method	Variable	Steering wheel angle		Lane position	
		F-value	p-value	F-value	p-value
cumulative	environment	F(1,704)=46.0266	< 0.0001	F(1,704)=17.9677	< 0.0001
	interface	F(2,704)=165.7317	< 0.0001	F(2,704)=223.7631	< 0.0001
	environment \times interface	F(2,704)=11.5925	< 0.0001	F(2,704)=5.2674	0.0054
per-glance	environment	F(1,676)=36.1101	< 0.0001	F(1,676)=74.1802	< 0.0001
	interface	F(2,676)=50.0705	< 0.0001	F(2,676)=47.592	< 0.0001
	environment \times interface	F(2,676)=3.7742	0.0234	F(2,676)=3.6415	0.0267

Table 4.4 Results of two-way ANOVAs for cumulative and per-glance cross-correlation results.

In order to examine at which levels of the *interface type* there exist significant differences between the two *environment types*, we proceeded with pairwise comparisons. Table 4.5 gives an overview of the magnitudes of the most prominent cumulative and per-glance cross-correlation peaks for both studies. The “comparison” column indicates which study produced a larger cross-correlation result for each *interface type*, while the “p-value” column indicates whether the comparison is significant or not

(p-values smaller than 0.05 are presented in bold face). As we can see both cumulative and per-glance steering wheel angle cross-correlation results are consistently larger in the city environment, while in case of lane position highway environment produced consistently larger results. Practically all comparisons indicated significant differences, except for cumulative lane position cross-correlation results for B condition ($p=0.868$) and per-glance steering wheel angle cross-correlation results for E condition ($p=0.1421$).

Cross-correlation method	Interface type	Steering wheel angle				Lane position			
		Highway	Comparison	City	p-value	Highway	Comparison	City	p-value
cumulative	B	3.647	<	7.714	0.0001	0.157	>	0.1539	0.868
	E	13.88	<	18.19	0.0371	0.4291	>	0.334	0.0208
	D	24.7	<	41.6	0.0001	0.7903	>	0.6182	0.0002
per-glance	B	0.9405	<	1.613	0.0001	0.04415	>	0.03267	0.0001
	E	1.798	<	2.096	0.1421	0.05438	>	0.04098	0.0001
	D	2.136	<	3.165	0.0001	0.07199	>	0.04914	0.0001

Table 4.5 Comparing magnitudes of most prominent cross-correlation peaks for two reference studies.

Based on the above results we can conclude that hypothesis H4 is supported, since our cross-correlation method managed to detect differences between the two driving environments.

By observing the magnitudes of the most prominent peaks in both cumulative and per-glance cross-correlation results, we can see that the ordering is the same in both driving environments: D condition resulted in the largest effect, followed by E and B conditions. However, it would also be interesting to compare relative sizes of the effects

of each *interface type* within each *environment type*. Table 4.6 shows the ratios of the magnitudes of the most prominent peaks in both cumulative and per-glance cross-correlation results between *interface types* for each *environment type*. If we compare the ratios between the highway and city environment we can see that they are mixed for both steering wheel angle and lane position cross-correlation results. However, the magnitudes of the ratios are fairly similar. This indicates that, even though the differences in absolute amplitudes do exist, the change in the environment did not affect highly the relative differences between interface types within each environment. Larger ratios observed for steering wheel angle compared to lane position can be explained by steering wheel angle's faster dynamics, as we discussed in Chapter 3.

Cross-correlation method	Ratio	Steering wheel angle			Lane position		
		Highway	Comparison	City	Highway	Comparison	City
cumulative	D/B	6.77	>	5.39	5.03	>	4.02
	D/E	1.78	<	2.29	1.84	<	1.85
	E/B	3.81	>	2.36	2.73	>	2.17
per-glance	D/B	2.27	>	1.96	1.63	>	1.5
	D/E	1.19	<	1.51	1.32	>	1.2
	E/B	1.91	>	1.3	1.23	<	1.25

Table 4.6 Relative differences between interface types in both environments.

4.2.4 Comparing Average-Based Driving Performance

Measures and Cumulative Cross-Correlation Results

As a conclusion to this section we would like to discuss how our cumulative cross-correlation results compare to average-based measures. This comparison is sound, since average-based measures also reflect the overall effects of in-vehicle interactions on driving. In both iPod studies we had a chance to see that at least some of the average-based driving performance measures reached significance. However, significant differences between interface types were detected more often in city driving. As we can see in Table 4.7 variances of steering wheel angle and velocity detected differences between all three interfaces (iPod interactions) in city driving. On the other hand, in highway driving variance of steering wheel angle was the only average-based performance measure which detected differences between all interface types. This finding can be explained by the increase in difficulty caused by the city environment, thus resulting in in-vehicle interactions producing larger effects that were successfully detected by the average-based measures. The ranking of measures that reached significance matches the ranking observed with our cumulative cross-correlation results, which provides clear support for construct validity. Nevertheless, there exist average-based measures which either did not reach significance (such as average velocity in highway driving and lane position variance in city driving) or did not detect differences between all interface types (lane position variance and velocity variance in highway driving and average velocity in city driving); however, the differences between all interface types were successfully detected using our cross-correlation method.

Study	Average-based measure	Main effect?	Pairwise comparisons		
			B - E	B - D	E - D
highway	lane position variance	YES	NO	NO	YES
	steering wh. angle variance	YES	YES	YES	YES
	velocity variance	YES	NO	NO	YES
	average velocity	NO			
city	lane position variance	NO			
	steering wh. angle variance	YES	YES	YES	YES
	velocity variance	YES	YES	YES	YES
	average velocity	YES	NO	NO	YES

Table 4.7 Comparison of significant effects detected using average-based measures for the two reference studies.

4.3 Obtaining Predictors of Cross-Correlation Results

Many of the conclusions from the previous section directly facilitate the process of explaining the underlying mechanisms of the cross-correlation method. Also, the way we conducted the two iPod studies (i.e., the reference experiments approach) and the highly controlled environments without (or at least minimized) confounding variables help significantly with drawing conclusions.

The purpose of this section is to propose a set of variables (predictors) which may have an important influence on the cross-correlation results. We expect that the same set of predictors will be revealed in both reference studies. Therefore, we will analyze both studies separately; however, we will use the same procedure. Furthermore, we will pool the data together in order to observe the effect of the driving environment as well.

Our method produces two types of cross-correlation results: cumulative and instance-based (per-glance in our case). Since our method relies on visual attention and

driving performance, we should devise a set of variables that describe both of these aspects well. In hypothesis H3 we proposed to examine the following variables: PDT away from the road, number of glances, glance duration and average absolute amount of change in lane position, steering wheel angle and vehicle heading. However, the predictors do not have to be the same for both types of cross-correlation results. The following paragraphs will provide explanations behind our decisions for the particular choice of variables.

4.3.1 Describing Visual Attention

Horrey et al. [15] found that the variability of lane position increased as the scanning of the outside world (PDT on outside world) decreased. This relationship suggests that PDT may be an important variable to consider in explaining the cumulative cross-correlation results. Namely, PDT describes the *overall* visual attention on an experimental segment. Similarly, cumulative cross-correlation describes the *overall* change in driving performance measures (in our case, lane position and steering wheel angle) influenced by *overall* visual attention over the same segment. Therefore, we can argue that they have the same “underlying” nature and that PDT may have a significant influence on the obtained results. In our analysis we propose to use PDT away from the road (as opposed to Horrey et al. who used PDT on the outside world), which can be obtained directly from PDT on the outside world as follows:

$$PDT_{away_from_road} = 100 - PDT_{outside_world}.$$

Alternatively, instead of using PDT away from the road, we can simultaneously use glance duration and number of glances directed away from the road.

These two variables provide low level description of the visual attention. However, when looking at the *overall* visual attention, knowing either PDT or glance duration + number of glances is sufficient. This can be demonstrated using a simple example. Let us assume that we have N glances, each p seconds long, on a t seconds long segment. Knowing all this we can calculate the PDT away from the road for that segment as:

$$PDT_{away_from_road} = \frac{N \cdot p}{t} \cdot 100\%.$$

In reality, of course, glances are not of equal duration. Nevertheless, the above example illustrates the approximate (at least asymptotical) equivalence of the information provided by PDT and glance duration + number of glances. Therefore, using all three variables concurrently in describing cumulative cross-correlation results would not provide any additional information and would likely cause multicollinearity.

If we look at instance-based (per-glance) cross-correlation result, it provides information about the change in driving performance influenced by individual instances of secondary task engagement. In our case those are individual glances directed away from the road. Neither PDT nor number of glances would be adequate variables for characterizing per-glance cross-correlation results, since they describe *overall* visual attention. However, glance duration describes *individual* glances (instances of secondary task engagement). Therefore, we will use glance duration in our models for characterizing per-glance cross-correlation results.

4.3.2 Describing Driving Performance

Since our method has been applied to lane position and steering wheel angle, it is logical to expect that these variables may be useful in characterizing our cross-correlation results. Therefore, we will use both of these variables in our analyses.

One additional variable which we expect may have an important influence is vehicle heading. Vehicle heading represents the angle between the tangential direction of the vehicle and north direction. Vehicle heading is measured in degrees and has positive values in the counter-clockwise direction and negative values in the clockwise direction. In our driving simulator, if the vehicle is perfectly aligned with the north direction, its heading equals 0° . The reason we believe that vehicle heading may be important is that drivers may decide to apply a different amount of change on the steering wheel depending on the heading of the vehicle. This is of course true for lane position as well. However, a driver can drive close to the edge of the road or the opposite lane indefinitely without the need to change the position of the vehicle. On the other hand, unsatisfactory vehicle heading (yaw too far to the left or right) may provide additional incentive for correcting the position in the lane.

A certain amount of redundancy can be expected between these three variables, since all of them are impacted by the changes on the common controller – steering wheel. However, cases exist when the information obtained by vehicle heading may complement the information obtained from steering wheel angle and lane position. Figure 4.19 illustrates this. Let us assume, without the loss of generality, that the road is perfectly aligned with the north direction. We can devise four possible cases:

- a) If the vehicle is parallel to the road and steering wheel angle is fixed at 0° : both lane position and vehicle heading are not changing (Figure 4.19, upper left).
- b) If the vehicle is not parallel to the road and steering wheel angle is fixed at 0° : lane position is changing, vehicle heading is not changing (Figure 4.19, upper right).
- c) If the steering wheel angle is fixed at some value different than 0° : both lane position and vehicle heading are changing (Figure 4.19, lower left).
- d) If the steering wheel angle is changing: both lane position and vehicle heading are changing (Figure 4.19, lower right).

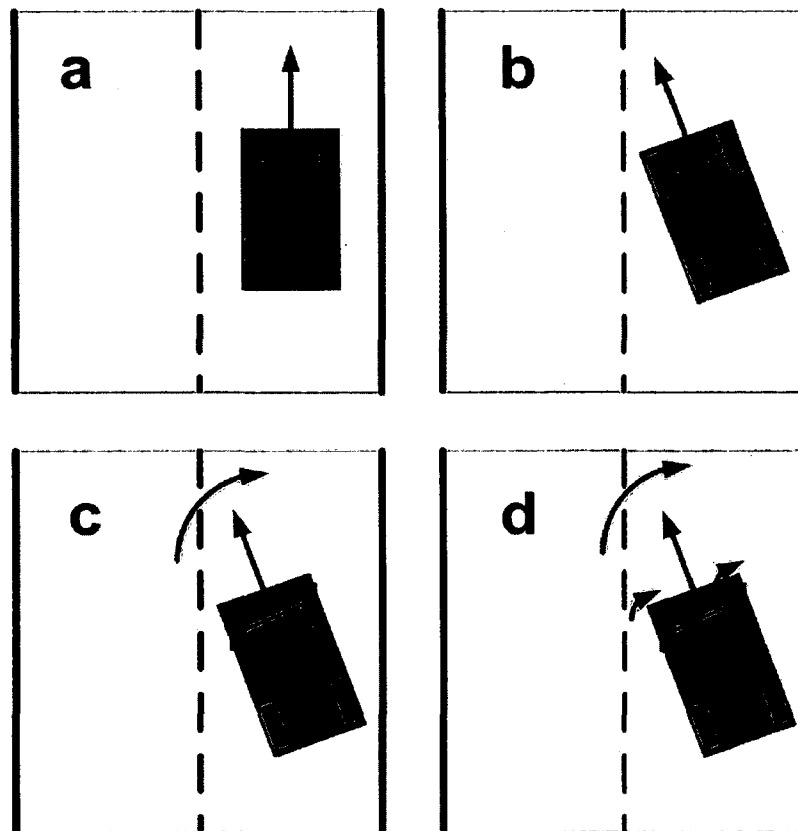


Figure 4.19 Illustration of the relationships between steering wheel angle, lane position and vehicle heading.

The examples presented in Figure 4.19 illustrate when vehicle heading may complement the information obtained from steering wheel angle and lane position. As we can see, cases a) and d) illustrate when all three variables behave in the same fashion: either not change or change together. However, in b) vehicle heading may provide additional information since it is not changing while lane position is. Similarly, in c) steering wheel angle is not changing while vehicle heading is. Based on these examples we decided to include vehicle heading in our analyses.

4.3.3 Data Collection

Regression analysis requires two types of variables: *independent* (or predictor) variables, which are used for predicting the response of a *dependent* variable. The following sections will describe both independent and dependent variables that will be used in our regression analyses.

Independent Variables

PDT away from the road: If N is the total number of samples in the current segment and P is the number of samples indicating looking away from the road, then PDT away from the road (PDT_AFR) can be calculated as follows:

$$PDT_AFR = \frac{P}{N} \cdot 100\%.$$

No transformation is necessary here, since PDT already provides a single value which characterizes the whole segment.

Number of Glances: As its name suggests, number of glances (NG) was obtained simply by counting the total number of glances directed off-road for each

segment. Therefore, no further transformation is necessary. Note that all the rules for filtering glances explained in Section 3.1.2 apply here as well.

Glance duration: Since in general multiple glances may occur on any segment, average value is the appropriate transformation to be used. If M is the total number of glances in the current segment and g_i is the duration of the i^{th} glance, then the average glance duration (AGD) can be obtained as follows:

$$AGD = \frac{1}{M} \cdot \sum_{i=1}^M g_i.$$

Driving performance: If we recall from Chapter 3, we transformed steering wheel angle and lane position using the absolute value of change (AVC) function before applying cross-correlation. To be consistent with this choice of transformation, we decided to apply the same type of transformation to all three driving performance measures (steering wheel angle, lane position and vehicle heading). However, in order to obtain a single descriptive value per segment for each driving variable, we also averaged the result obtained from AVC. We will refer to this transformation as the average absolute value of change (AAVC) and for an arbitrary sequence x it is defined as follows:

$$AAVC\{x\} = \frac{1}{N-1} \cdot \sum_{i=2}^N \frac{|x(n) - x(n-1)|}{T_s},$$

where N is the length of sequence x and T_s is the sampling period. This particular transformation is similar in nature to standard deviation, which may be used as a possible alternative.

Dependent Variables

Since we intend to provide explanations for both cumulative and per-glance cross-correlation results, each approach provides one corresponding dependent variable. We will refer to the variable which holds the cumulative result as XCORR_CML_X and the per-glance result as XCORR_PG_X, where “X” refers to either steering wheel angle (SWA) or lane position (LP).

Since cross-correlation functions (both cumulative and per-glance) represent time series data, a suitable transformation is required in order to obtain a unique value which characterizes each experimental segment. Inspired by our “area below the curves” approach presented in Section 3.1.5, we decided to calculate the areas below the cross-correlation functions for each segment. Prominent peaks provide a convenient visual representation of the cross-correlation results through their magnitude and time lag. However, there are two reasons we decided to use areas rather than magnitudes of the peaks in our regression models. First, areas consider a wider time interval after the gaze returns back to the road, while peaks consider only individual instants in time. Since we had a chance to see that there is usually a range of lags (around the most prominent peak) for which the cross-correlation functions are significant (larger than the $p=0.05$ level), we can say that areas can extract more information about the changes in driving performance over time. And second, the most prominent peaks in our cross-correlation results represent average changes in our driving performance measures which we obtained over a number of experimental segments. However, the individual peaks do not have to occur at exactly the same location as the most prominent peak. Therefore, areas can account for these differences in individual segments. We also have to notice that both prominent peaks and areas always revealed the same ranking of the interaction types in all previous

studies, which indicates that both approaches provide the same conclusions. For each segment we decided to calculate areas below the cross-correlation functions between the lags of 0 and 1 second. The decision to consider this interval is supported by the fact that the most prominent peaks for both lane position and steering wheel angle cross-correlation functions on average occurred around the lag of 0.6 seconds.

Summary of Selected Variables

Before we delve into the data preparation for regression analyses, we summarize the above independent variables in Table 4.8.

Variable name	Transformation	Abbreviation	Cross-correlation function	
			Cumulative	Per-glance
PDT Away From Road	-	PDT_AFR	✓	
Number of Glances	-	NG	✓	
Glance Duration	average	AGD	✓	✓
Steering Wheel Angle	AAVC	AAVC_SWA	✓	✓
Lane Position	AAVC	AAVC_LP	✓	✓
Vehicle Heading	AAVC	AAVC_VH	✓	✓

Table 4.8 Overview of the proposed independent variables and their corresponding transformations used in regression analyses.

The check marks in Table 4.8 indicate which variables are used in predicting the results for each cross-correlation method (cumulative and per-glance). As stated before, for cumulative cross-correlation results, either PDT or glance duration + number of glances should be used in the regression models, but not both of those concurrently. The “transformation” column indicates a specific transformation function which was

applied to each variable. The “abbreviation” column gives the short names of the final (transformed) variables as will be used in the regression models.

Each of the above variables was obtained for each experimental segment. Since each of 12 participants completed 10 experimental segments, this amounts to a total of 120 potential segments that are available for the analyses of each experimental condition (B, E and D). Therefore, every segment s_i ($i = 1, \dots, 120$) can be characterized with a set of values, one for each of the above variables:

$$s_i \rightarrow \{PDT_AFR_i, NG_i, AGD_i, AAVC_SWA_i, AAVC_LP_i, AAVC_VH_i\}.$$

Some variables, such as PDT and number of glances, by definition provide only a single value for each segment. However, driving performance variables, for example, are time series data, which cannot be used directly. Therefore, appropriate transformation had to be applied first in order to obtain a single-value description of an experimental segment. Note that the same set of variables and the corresponding procedure were used in both reference studies.

4.3.4 Data Conditioning for Regression Analysis

There are two main steps that we performed in conditioning the data for regression analysis: normalizing distributions of variables and handling outliers.

Normalizing Distributions of Variables

Regression analysis does not require independent variables to be normally distributed. Nevertheless, skewed distributions often cause statistical problems, such as heteroscedasticity and influence [132]. Therefore, we decided to apply appropriate

transformations to each variable in order to bring their distributions as close as possible to “normal” looking.

All of the variables we are using in this analysis are positively skewed, which means that they have a long upper tail. Skewed distributions can often be “normalized” by applying power transformations [132]. There are at least two advantages to using power transformations: they make skewed distributions more symmetrical and also may pull in outliers. Let us say that X is our independent variable, q is the power exponent and $X' = X^q$ is the transformed original variable. Depending on the exponent q we can obtain different effects: $q > 1$ reduces negative skew by shifting the weight to the upper tail, while $q < 1$ reduces positive skew by pulling in the upper tail. Since we are dealing with positively skewed variables, $q < 1$ was the appropriate choice in each case. We have to note that logarithmic transformation ($\log(X)$) is also very commonly used. However, in our case it was too powerful and often resulted in shifting the distributional shape from positive to negative skew. Therefore, we used power transformations only.

As suggested in [132], we can judge how close to normal a symmetrical distribution is, by comparing its standard deviation with $IQR/1.35$. IQR represents the interquartile range and is calculated as the difference between the third and first quartile. If the standard deviation of the given variable is similar to $IQR/1.35$ we can say that its distribution has tails which are close to normal. We used this as a benchmark to judge which power transformation provided an acceptable result for each variable.

Table 4.9 outlines the exponents (q) of power transformations (X^q) that have been applied to each variable (X) in both reference studies for cumulative and per-glance cross-correlation results.

Variable name	Abbreviation	Cumulative (CML)		Per-glance (PG)	
		City	Highway	City	Highway
PDT Away From Road	PDT_AFR	0.5	0.4	-	-
Glance Duration	AGD	0.9	0.8	0.6	0.2
Number of Glances	NG	0.8	0.7	-	-
Steering Wheel Angle	AAVC_SWA	0.4	0.2	0.5	0.3
Lane Position	AAVC_LP	0.2	0.3	0.2	0.3
Vehicle Heading	AAVC_VH	0.4	0.2	0.4	0.2
Steering wheel angle cross-correlation	XCORR_(CML or PG)_SWA	0.4	0.3	0.5	0.4
Lane position cross-correlation	XCORR_(CML or PG)_LP	0.5	0.4	0.5	0.6

Table 4.9 Exponents of power transformations used for normalizing the data.

Missing exponents in Table 4.9 indicate that the corresponding variables were not used in creating the regression model for that particular cross-correlation result. The exponents differ between the variables, because their distributions had different levels of skew. Note that the transformations applied to common independent variables between the “cumulative” and “per-glance” columns do not have to be exactly the same, because cumulative and per-glance cross-correlation results may not use the same experimental segments in their calculations. Specifically, since per-glance cross-correlation results provide the amount of change in a driving performance variable introduced by individual glances, segments with no off-road glances are not used in the calculations. On the other hand, cumulative cross-correlation results include all segments in the calculations, since they provide an overall response. Nevertheless, we can see that the exponents are mostly similar between the two.

Handling Outliers

As we mentioned in the previous subsection, power transformations can also pull in outliers. Nevertheless, we still had to check for the existence of outliers that remained after applying the transformations. Since we proposed to use multiple variables in devising our regression models, we have to look at the outliers from the multivariate perspective. In other words, a data point may not be an outlier when looked at from the univariate standpoint, but can be an outlier when looked at from the multivariate standpoint. One widely used method for multivariate outlier detection is Mahalanobis distance [133]. This method identifies unusual data points that lie far from the multivariate center of the data, which we subsequently rejected as outliers. Using JMP 9.0 we applied this method to all of our variables.

Observing Distributions of the Transformed Variables

Figures 4.20 and 4.24 depict distributions of independent variables which were used in modeling the cumulative cross-correlation results for highway and city study, respectively. Distributions of their respective dependent variables are presented in Figures 4.21 and 4.25.

Similarly, Figures 4.22 and 4.26 present distributions of independent variables used in modeling per-glance cross-correlation results for highway and city study, respectively. Their corresponding dependent variables are depicted in Figures 4.23 and 4.27.

Each figure is organized as a table with cells showing distributions of individual variables. Graphs marked with “O” represent original data, while “T” indicates

transformed and outlier-free data. Transformed data is used in the next section for creating regression models.

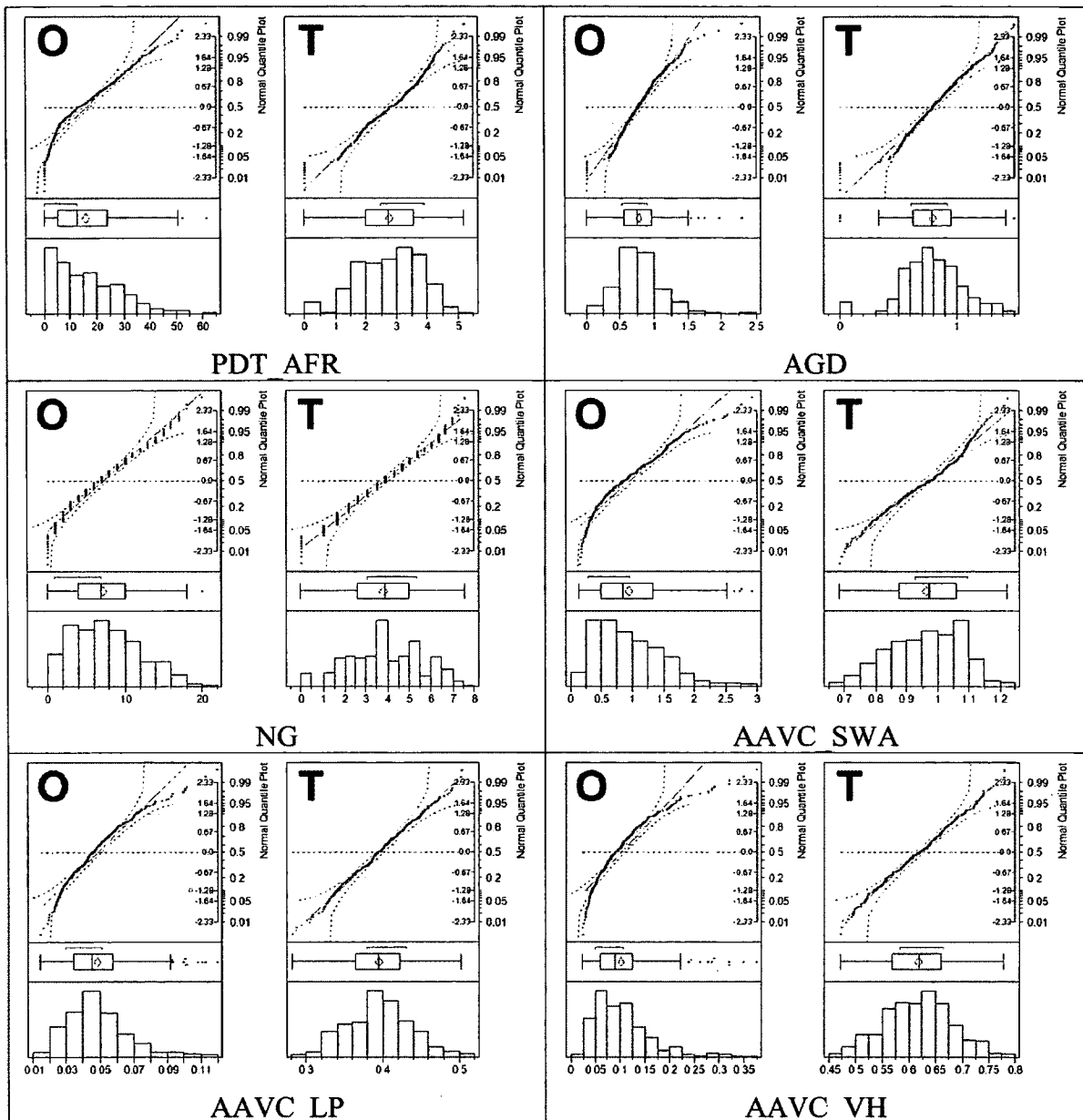


Figure 4.20 Highway study: distributions of independent variables used for modeling cumulative cross-correlation results.

Each graph contains three plots: normal quantile plot (top), box plot (middle) and histogram (bottom). If a distribution is close to normal, we expect its histogram and

box plot to be approximately symmetric and the normal quantile plot to approximately follow a straight line (secondary diagonal line in each graph). As we can see, power transformations considerably improved normality of all variables.

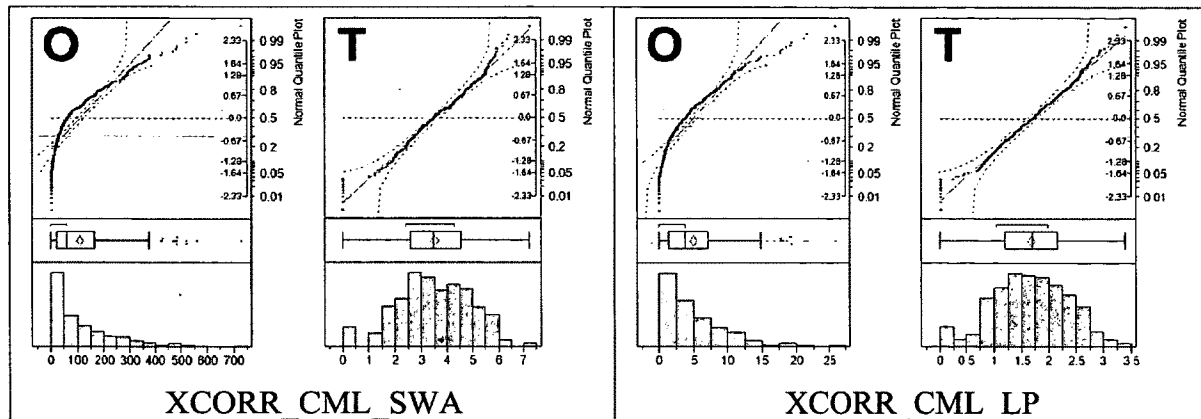


Figure 4.21 Highway study: distributions of dependent variables used for modeling cumulative cross-correlation results.

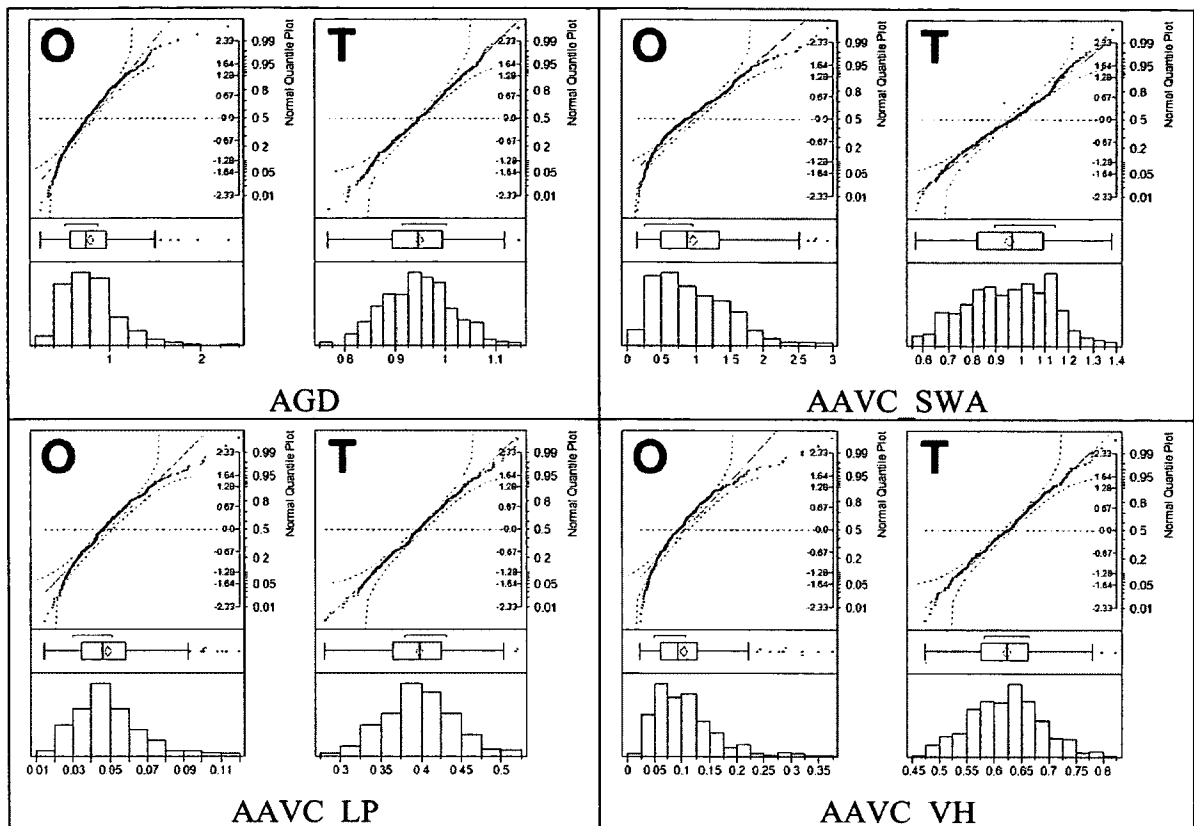


Figure 4.22 Highway study: distributions of independent variables used for modeling per-glance cross-correlation results.

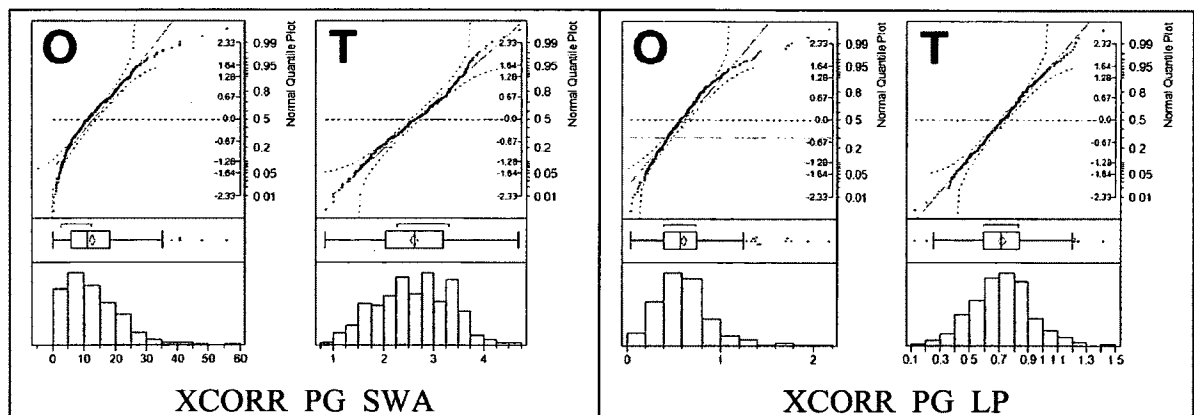


Figure 4.23 Highway study: distributions of dependent variables used for modeling per-glance cross-correlation results.

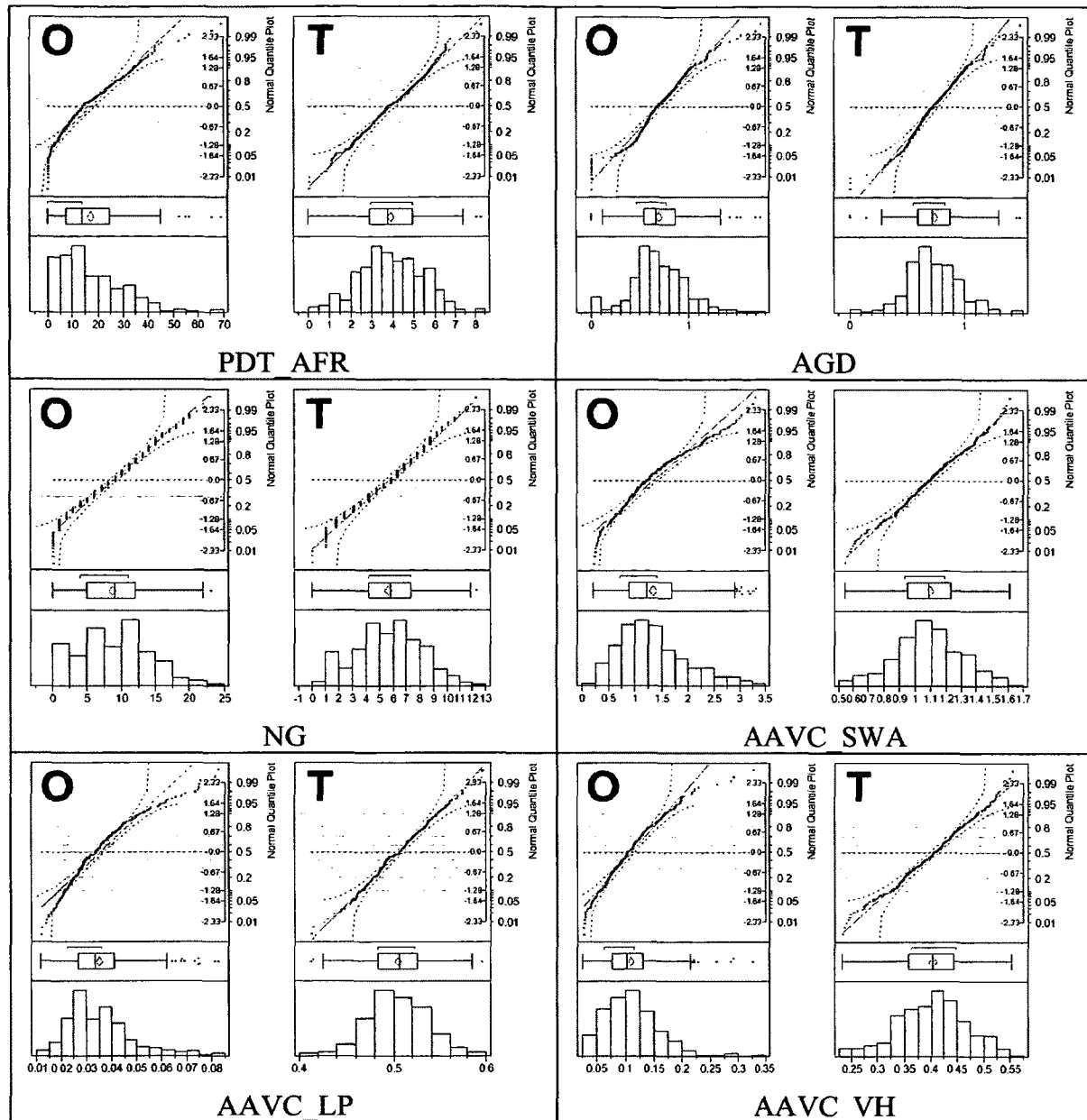


Figure 4.24 City study: distributions of independent variables used for modeling cumulative cross-correlation results.

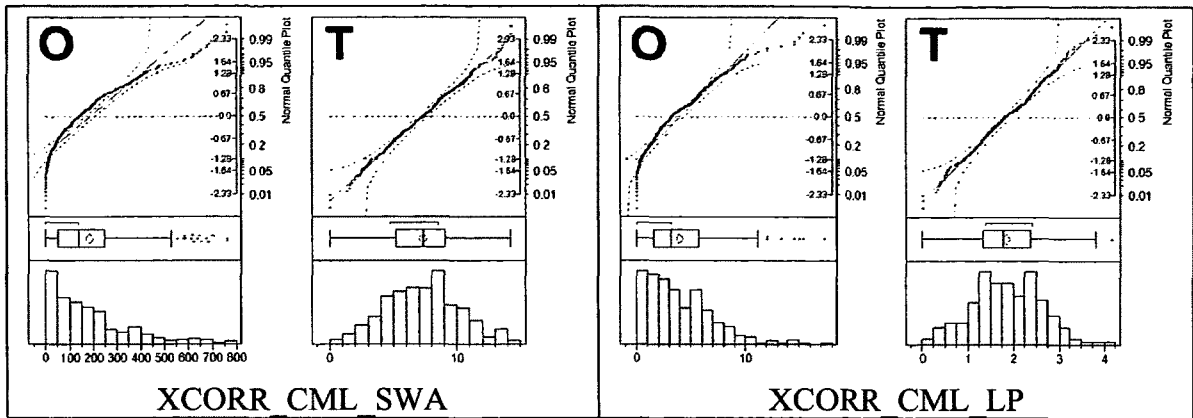


Figure 4.25 City study: distributions of dependent variables used for modeling
cumulative cross-correlation results.

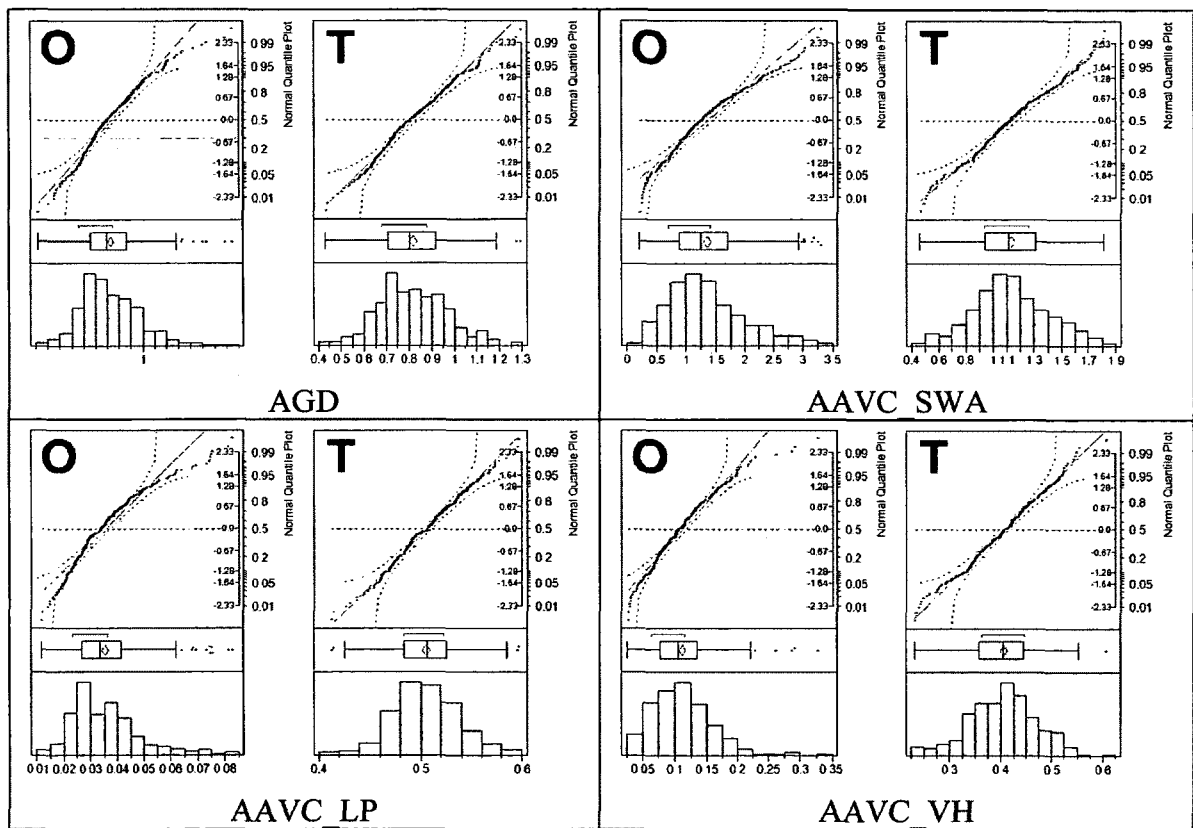


Figure 4.26 City study: distributions of independent variables used for modeling per-
glance cross-correlation results.

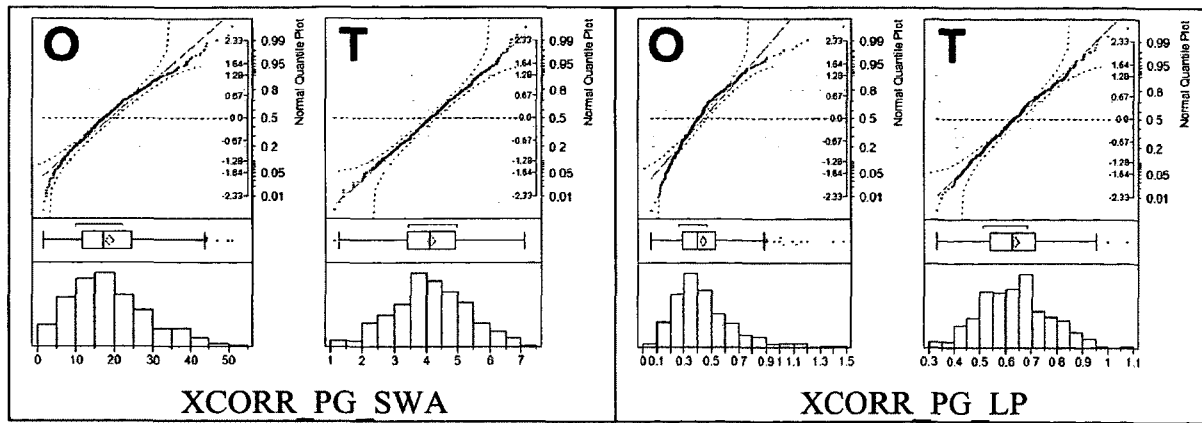


Figure 4.27 City study: distributions of dependent variables used for modeling per-glance cross-correlation results.

4.3.5 Creating Regression Models for Reference Studies

This section presents the regression models which were created using the transformed variables described in the previous section. There are eight regression models in total, one for each combination of environment (highway, city), cross-correlation result (cumulative, per-glance) and driving performance variable (steering wheel angle, lane position).

Regression models strip away the random errors or noise thus revealing the underlying relationship between regressors (independent variables) and the response (cross-correlation results). As suggested in Section 4.3.1 (pg. 215), we can use either PDT_AFR or AGD + NG in modeling cumulative cross-correlation results. The first model (we will refer to it as CML_M1_X, where “X” represents SWA or LP) uses the following regression equations for modeling the cumulative cross-correlation results:

$$XCORR_CML_SWA = a_0 + a_1 \cdot PDT_AFR + a_2 \cdot AAVC_SWA + a_3 \cdot AAVC_VH$$

$$XCORR_CML_LP = b_0 + b_1 \cdot PDT_AFR + b_2 \cdot AAVC_LP + b_3 \cdot AAVC_VH$$

Equation 4.1 Modeling cumulative cross-correlation results using CML_M1_X model.

The alternative model using AGD and NG (we will refer to it as CML_M2_X) can be defined as follows:

$$XCORR_CML_SWA = c_0 + c_1 \cdot AGD + c_2 \cdot NG + c_3 \cdot AAVC_SWA + c_4 \cdot AAVC_VH$$

$$XCORR_CML_LP = d_0 + d_1 \cdot AGD + d_2 \cdot NG + d_3 \cdot AAVC_LP + d_4 \cdot AAVC_VH$$

Equation 4.2 Modeling cumulative cross-correlation results using CML_M2_X model.

Similarly, the regression equations used for modeling the per-glance cross-correlation results (we will refer to this model as PG_M_X) are as follows:

$$XCORR_PG_SWA = e_0 + e_1 \cdot AGD + e_2 \cdot AAVC_SWA + e_3 \cdot AAVC_VH$$

$$XCORR_PG_LP = f_0 + f_1 \cdot AGD + f_2 \cdot AAVC_LP + f_3 \cdot AAVC_VH$$

Equation 4.3 Modeling per-glance cross-correlation results using PG_M_X model.

We expect that all of the above variables will positively contribute to the cross-correlation results (both cumulative and per-glance). In other words, as the values of these variables increase, the cross-correlation results are expected to increase as well. Therefore, we expect that all of the above coefficients associated with these variables should result with positive signs.

All regression analyses were performed in JMP 9.0 using the following steps for each cross-correlation model (cumulative and per-glance) and each study (city and highway):

1. Fit the largest possible regression model using all of the variables proposed in the above equations.
2. This step is concerned with checking the influence of the individual observations (data samples) on the fitted regression models and consists of multiple steps which should be observed in concert:
 - a. Check for high leverage points using hat diagonals. Hat diagonals determine the amount of weight each observation has on its own prediction. A high leverage value close to 1 indicates that an observation entirely predicts itself,

which is not desirable in general. Although high leverage points have the potential to bias or distort the regression model estimates, they are not necessarily bad data points. If the high leverage points are valid data points, their presence can actually improve the regression model. However, if a high leverage point is an outlier, the fitted model may be biased in its predictions. The existence of outliers is checked by observing the distribution of the studentized residuals presented in step 3c. High leverage points can also be influential, which is checked with the Cook's D statistic in step 3b.

- b. Use Cook's D statistic to identify influential points. Cook's D value is calculated for each observation point. It measures how much the model coefficient estimates would change if the i^{th} observation were to be removed from the dataset. The higher the Cook's D value, the higher the influence. Values above 1 indicate some influence, while values in 3 and 4 digits indicate extreme influence.
- c. Check for outliers by observing the distribution of the studentized residuals. Values far from ± 3 indicate potential outliers.
- d. If all of the data samples have the above statistics within the proposed limits, we can conclude that no obvious problems with the dataset exist and we can continue with the process of obtaining the best regression model. If a data sample is a high leverage point (based on hat diagonals) and an outlier (based on studentized residuals), we remove it from the dataset. If a data sample is an influential point (based on Cook's D), we have to check both studentized residuals and the raw data and if it proves to be an outlier we remove it from

the dataset. After excluding the above data samples, we fit the regression models again using the remaining data samples.

3. Examine the significance of each regression coefficient ($a_i, b_i, c_i, d_i, e_i, f_i$) and keep the variables whose coefficients are significant at the $p=0.05$ level in the model. If any of the variables should be removed from the model, we fit the model again (with the smaller subset of variables) in order to obtain new coefficient estimates.
4. Check for normality of residuals by plotting the normal quantile plot for the studentized residuals. If the plot approximates a straight line, we can conclude that the normal assumption is satisfied. This is the most important assumption for the regression analysis.
5. Check for heteroscedasticity by plotting residuals versus predicted values. If the points presented in this graph are approximately randomly dispersed around 0 (on vertical axis), it indicates that heteroscedasticity is not an issue. Furthermore, the lack of nonrandom patterns indicates that the model has no missing variables.
6. Check for multicollinearity among the regressors in the model by observing the variance inflation factor (VIF) statistic. VIF measures how much the variance of a coefficient estimate is inflated by multicollinearity. VIF values larger than 30 indicate considerable multicollinearity, while values in 3 and 4 digits indicate severe multicollinearity.
7. Finally, in reporting the results of each regression analysis, we will use the following:

- a) Adjusted coefficient of determination (R^2), which represents the proportion of variance in the dependent variable that is explained by the regression model.
- b) Coefficient estimates ($a_i, b_i, c_i, d_i, e_i, f_i$) in our regression equations.
- c) p-values, which indicate whether a particular coefficient is statistically significant (in our case, we use $p < 0.05$).
- d) Standardized betas, which represent the coefficient estimates that would have been obtained from the regression if all the variables were standardized to a mean of 0 and variance of 1 [132]. As such, they show how many standard deviations a dependent variable would change per 1 standard deviation change in a particular independent variable (everything else being equal). Thus, they are often used in multiple regression analyses to determine which variables have a higher effect on the dependent variable, when the variables have different units of measurement.
- e) VIF values for checking multicollinearity.

The above procedure did not reveal any problems with the datasets used for modeling any of our cross-correlation results. No high leverage or influential points have been observed. The distributions of studentized residuals resembled normal distribution and residuals versus predicted plots indicated no obvious problems with heteroscedasticity and missing variables. All VIF values are much smaller than 30, indicating no issues with multicollinearity. In order to make the following sections easier to read, we include the graphs for studentized residuals and residuals versus predicted plots in Appendix C. The following sections outline the results of the regression analyses.

Modeling Cumulative Cross-Correlation Results in Highway Study

Cumulative steering wheel angle cross-correlation models: The following models for cumulative steering wheel angle cross-correlation results are analyzed here (CML_M1_SWA and CML_M2_SWA):

$$XCORR_CML_SWA = a_0 + a_1 \cdot PDT_AFR + a_2 \cdot AAVC_SWA + a_3 \cdot AAVC_VH$$

$$XCORR_CML_SWA = c_0 + c_1 \cdot AGD + c_2 \cdot NG + c_3 \cdot AAVC_SWA + c_4 \cdot AAVC_VH$$

Table 4.10 shows the results of the analysis for CML_M1_SWA model. The model provided a very good fit, with approximately 91% of variance explained. All coefficient estimates are positive and significant (significant coefficients will be presented in bold face in all tables).

CML_M1_SWA	R² = 0.91		Highway		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	a_0	-4.6048	< 0.0001	-	-
PDT_AFR	a_1	0.8923	< 0.0001	0.6644	1.4193
AAVC_SWA	a_2	4.3857	< 0.0001	0.3642	5.0643
AAVC_VH	a_3	2.347	0.0107	0.1056	5.9073

Table 4.10 Highway study: Coefficient estimates for CML_M1_SWA model.

Similarly, Table 4.11 shows the results for CML_M2_SWA model. The model explained about 93% of variance and all coefficient estimates are positive and significant.

<i>CML_M2_SWA</i>	R² = 0.93		Highway		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	c_0	-4.0305	< 0.0001	-	-
AGD	c_1	0.778	< 0.0001	0.1472	1.5545
NG	c_2	0.4932	< 0.0001	0.6084	2.0849
AAVC_SWA	c_3	4.0823	< 0.0001	0.339	5.0713
AAVC_VH	c_4	1.8279	0.0213	0.0822	5.9129

Table 4.11 Highway study: Coefficient estimates for CML_M2_SWA model.

Cumulative lane position cross-correlation models: The following models for cumulative lane position cross-correlation results are analyzed here (CML_M1_LP and CML_M2_LP):

$$XCORR_CML_LP = b_0 + b_1 \cdot PDT_AFR + b_2 \cdot AAVC_LP + b_3 \cdot AAVC_VH$$

$$XCORR_CML_LP = d_0 + d_1 \cdot AGD + d_2 \cdot NG + d_3 \cdot AAVC_LP + d_4 \cdot AAVC_VH$$

Table 4.12 shows the results of the analysis for CML_M1_LP model. We can see that the model explained about 89% of variance. All coefficients are significant and positive.

<i>CML_M1_LP</i>	R² = 0.89		Highway		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	b_0	-2.1901	< 0.0001	-	-
PDT_AFR	b_1	0.4644	< 0.0001	0.7021	1.3585
AAVC_LP	b_2	4.3136	< 0.0001	0.2636	2.8828
AAVC_VH	b_3	1.4266	< 0.0001	0.1303	3.1651

Table 4.12 Highway study: Coefficient estimates for CML_M1_LP model.

Finally, Table 4.13 shows the results for the CML_M2_LP model. The model explained about 93% of variance. All coefficients are statistically significant except AAVC_VH ($p=0.3468$). As indicated in the general analysis procedure at the beginning of this section, if a coefficient proves to be non-significant, we remove it from the model and perform another regression analysis without it. Therefore, Table 4.13 shows the results as obtained *without* AAVC_VH in the model. However, we kept AAVC_VH in the table for completeness in order to emphasize that it was initially used in the model, but it did not prove to be significant. This same principle will be applied each time a coefficient proves to be non-significant.

<i>CML_M2_LP</i>	$R^2 = 0.93$		Highway		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	d_0	-1.7107	< 0.0001	-	-
AGD	d_1	0.2285	< 0.0001	0.0878	1.5119
NG	d_2	0.2862	< 0.0001	0.7169	1.7525
AAVC_LP	d_3	5.3483	< 0.0001	0.3269	1.2569
AAVC_VH	d_4	0.282	0.3468		

Table 4.13 Highway study: Coefficient estimates for CML_M2_LP model.

Modeling Cumulative Cross-Correlation Results in City Study

We will follow the same procedure as in the previous section in creating the regression models for the city study.

Cumulative steering wheel angle cross-correlation models: The following models for cumulative steering wheel angle cross-correlation results are analyzed here (CML_M1_SWA and CML_M2_SWA):

$$XCORR_CML_SWA = a_0 + a_1 \cdot PDT_AFR + a_2 \cdot AAVC_SWA + a_3 \cdot AAVC_VH$$

$$XCORR_CML_SWA = c_0 + c_1 \cdot AGD + c_2 \cdot NG + c_3 \cdot AAVC_SWA + c_4 \cdot AAVC_VH$$

Table 4.14 shows the results of the regression analysis for the CML_M1_SWA model. The model provided a very good fit, with approximately 91% of variance explained. All coefficients are statistically significant except for AAVC_VH (p=0.4865).

<i>CML_M1_SWA</i>	R² = 0.91		City		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	a_0	-5.2313	< 0.0001	-	-
PDT_AFR	a_1	1.1432	< 0.0001	0.5874	1.208
AAVC_SWA	a_2	7.3513	< 0.0001	0.5485	1.208
AAVC_VH	a_3	-0.9344	0.4865		

Table 4.14 City study: Coefficient estimates for CML_M1_SWA model.

Table 4.15 shows the results for CML_M2_SWA model. The model explained about 94% of variance and all coefficient estimates are positive and significant.

CML_M2_SWA	R² = 0.94		City		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	c_0	-5.1136	< 0.0001	-	-
AGD	c_1	1.3286	< 0.0001	0.1011	1.3062
NG	c_2	0.6791	< 0.0001	0.5815	1.5644
AAVC_SWA	c_3	5.9692	< 0.0001	0.4453	3.8924
AAVC_VH	c_4	2.6543	0.0275	0.0593	3.565

Table 4.15 City study: Coefficient estimates for CML_M2_SWA model.

Cumulative lane position cross-correlation models: The models for cumulative lane position cross-correlation results, CML_M1_LP and CML_M2_LP, are as follows:

$$XCORR_CML_LP = b_0 + b_1 \cdot PDT_AFR + b_2 \cdot AAVC_LP + b_3 \cdot AAVC_VH$$

$$XCORR_CML_LP = d_0 + d_1 \cdot AGD + d_2 \cdot NG + d_3 \cdot AAVC_LP + d_4 \cdot AAVC_VH$$

Table 4.16 shows the results of the regression analysis for the CML_M1_LP model. The model explained about 85% of variance. All coefficients are statistically significant and positive.

CML_M1_LP	R² = 0.85		City		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	b_0	-3.157	< 0.0001	-	-
PDT_AFR	b_1	0.3479	< 0.0001	0.6965	1.2092
AAVC_LP	b_2	5.4499	< 0.0001	0.2334	2.1357
AAVC_VH	b_3	2.1844	< 0.0001	0.1901	2.3235

Table 4.16 City study: Coefficient estimates for CML_M1_LP model.

Finally, Table 4.17 shows the regression results for the second model, CML_M2_LP. This model provided a better fit by explaining about 90% of variance. As before, all estimated coefficients are statistically significant and positive.

<i>CML_M2_LP</i>	R² = 0.90		City		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	d_0	-3.9457	< 0.0001	-	-
AGD	d_1	0.2072	0.0019	0.0614	1.2898
NG	d_2	0.215	< 0.0001	0.7173	1.3963
AAVC_LP	d_3	7.9029	< 0.0001	0.3384	2.3387
AAVC_VH	d_4	1.0684	0.0007	0.093	2.4982

Table 4.17 City study: Coefficient estimates for CML_M2_LP model.

Modeling Per-glance Cross-Correlation Results in Highway Study

Per-glance steering wheel angle cross-correlation model: This section analyzes PG_M_SWA model for per-glance steering wheel angle cross-correlation results:

$$XCORR_PG_SWA = e_0 + e_1 \cdot AGD + e_2 \cdot AAVC_SWA + e_3 \cdot AAVC_VH$$

Table 4.18 shows the results of the regression analysis for PG_M_SWA model. The model explained about 81% of variance. All coefficients are positive and highly significant.

<i>PG_M_SWA</i>	R² = 0.81		Highway		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	e_0	-3.2091	< 0.0001	-	-
AGD	e_1	2.0241	< 0.0001	0.188	1.0645
AAVC_SWA	e_2	2.9781	< 0.0001	0.7164	5.4679
AAVC_VH	e_3	1.7171	0.0071	0.1548	5.6156

Table 4.18 Highway study: Coefficient estimates for PG_M_SWA model.

Per-glance lane position cross-correlation model: This section analyzes the per-glance lane position cross-correlation model (PG_M_LP):

$$XCORR_PG_LP = f_0 + f_1 \cdot AGD + f_2 \cdot AAVC_LP + f_3 \cdot AAVC_VH$$

Table 4.19 presents the results of the regression analysis for the PG_M_LP model. As we can see, the model explained about 67% of variance and all coefficients are positive and significant.

<i>PG_M_LP</i>	$R^2 = 0.67$		Highway		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	f_0	-1.0856	< 0.0001	-	-
AGD	f_1	0.3481	0.0004	0.1178	1.0432
AAVC_LP	f_2	3.0039	< 0.0001	0.6613	3.0327
AAVC_VH	f_3	0.4516	0.0088	0.1484	3.0463

Table 4.19 Highway study: Coefficient estimates for PG_M_LP model.

Modeling Per-glance Cross-Correlation Results in City Study

Per-glance steering wheel angle cross-correlation model: This section analyzes the model for per-glance steering wheel angle cross-correlation results (PG_M_SWA):

$$XCORR_PG_SWA = e_0 + e_1 \cdot AGD + e_2 \cdot AAVC_SWA + e_3 \cdot AAVC_VH$$

The results of the regression analysis for the PG_M_SWA model are presented in Table 4.20. The model explained about 81% of variance, with all coefficients being statistically significant and positive.

<i>PG_M_SWA</i>	R² = 0.81		City		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	e_0	-1.2638	< 0.0001	-	-
AGD	e_1	1.0872	< 0.0001	0.1403	1.1292
AAVC_SWA	e_2	3.2654	< 0.0001	0.7805	3.3345
AAVC_VH	e_3	2.0054	0.0157	0.1127	3.5249

Table 4.20 City study: Coefficient estimates for PG_M_SWA model.

Per-glance lane position cross-correlation model: The model for per-glance lane position cross-correlation results (PG_M_LP) is as follows:

$$XCORR_PG_LP = f_0 + f_1 \cdot AGD + f_2 \cdot AAVC_LP + f_3 \cdot AAVC_VH$$

Table 4.21 shows the results of the regression analysis for the PG_M_LP model. As we can see, the model explained about 69% of variance. All estimated coefficients are statistically significant and positive.

<i>PG_M_LP</i>	$R^2 = 0.69$		City		
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	f_0	-0.9969	< 0.0001	-	-
AGD	f_1	0.0802	0.0073	0.0903	1.1509
AAVC_LP	f_2	2.8738	< 0.0001	0.6921	2.3458
AAVC_VH	f_3	0.28	0.003	0.1372	2.1666

Table 4.21 City study: Coefficient estimates for PG_M_LP model.

Discussion of the Regression Results for Individual Reference Studies

We started this section with the descriptions of the proposed regression models which we applied to both cumulative and per-glance cross-correlation results for two reference studies: highway and city driving. Based on the results presented in the previous subsections we can draw a general conclusion that our hypothesis H3 is supported for three reasons: first, all of the proposed variables proved to be statistically significant (except AAVC_VH in case of CML_M1_SWA model in city driving and CML_M2_LP model in highway driving), second, all of their corresponding coefficients demonstrated positive signs, and third, both steering wheel angle and lane position cross-correlation results can be described using the same variables. This indicates that the cross-correlation results indeed increase as the values of our proposed variables increase. This is an important result, because it suggests that interactions with in-vehicle devices should be performed in a way which minimizes the effects on these variables.

There are two models that we proposed for describing the cumulative cross-correlation results. One uses PDT_AFR (CML_M1_X model), while the second one uses AGD + NG (CML_M2_X model) for describing visual attention. Table 4.22 gives an overview of the standardized beta coefficients as well as the coefficients of determination calculated for the first model. Similarly, Table 4.23 summarizes the same results for the second model. The reason we are presenting standardized instead of actual coefficients in these tables is that they allow comparing the sizes of the effects of the independent variables on the dependent variable. Empty cells indicate that a variable was not used in the corresponding model, while “ns” indicates that a coefficient was not statistically significant. All other coefficients were statistically significant (with $p < 0.05$) and positive.

model	road	standardized beta coefficients				R ²
		PDT_AFR	AAVC_SWA	AAVC_LP	AAVC_VH	
CML_M1_SWA	highway	0.6644	0.3642		0.1056	0.91
	city	0.5874	0.5485		ns	0.91
CML_M1_LP	highway	0.7021		0.2636	0.1303	0.89
	city	0.6965		0.2334	0.1901	0.85

Table 4.22 Standardized beta coefficients for individual reference studies for model

CML_M1_X.

model	road	standardized beta coefficients					R ²
		NG	AGD	AAVC_SWA	AAVC_LP	AAVC_VH	
CML_M2_SWA	highway	0.6084	0.1472	0.339		0.0822	0.93
	city	0.5815	0.1011	0.4453		0.0593	0.94
CML_M2_LP	highway	0.7169	0.0878		0.3269	ns	0.93
	city	0.7173	0.0614		0.3384	0.093	0.9

Table 4.23 Standardized beta coefficients for individual reference studies for model

CML_M2_X.

If we compare the coefficients of determination in Tables 4.22 and 4.23 we can see that the second model (CML_M2_X) provides a better fit to the data. By just taking the average R^2 for each table (including both steering wheel angle and lane position), we can see that the first model (CML_M1_X) explains about 89% of variance, while the second model (CML_M2_X) explains about 92.5% of variance. We have to note here a well known fact that adding more variables to the regression model by definition increases R^2 . However, the difference in the number of variables between the two models is only one and from our experience adding an extraneous variable which accounts for a

trivial amount of variance increases R^2 only slightly. In our case, the observed increase of 3.5% between the two models can be attributed to the second model providing a better explanation of the visual attention (since the variables explaining driving performance remained the same between the two models). Namely, using both the average glance duration and number of glances provides more information about drivers' visual attention than by just looking at the PDT off-road. Nevertheless, both models explained a considerable amount of variance ($\geq 85\%$), which confirms that the variables selected in either case provide a very good explanation of the cumulative cross-correlation results.

We can also compare how well each model explains individual cumulative cross-correlation results for steering wheel angle and lane position. If we look at the coefficients of determination, we can see that the cumulative cross-correlation results pertaining to the steering wheel angle obtained somewhat better fits for both models. For the first model (CML_M1_X) the average R^2 for steering wheel angle is 0.91, while for lane position is 0.87. Similarly, for the second model (CML_M2_X) the average R^2 for steering wheel angle is 0.935, while for lane position is 0.915. This can be explained by the steering wheel angle being more sensitive to impacts of in-vehicle interactions, resulting from its faster dynamics (as we discussed in Chapter 3). Another support for this assertion comes from our cross-correlation functions which show much more pronounced peaks in case of steering wheel angle compared to lane position.

Regarding the size of the effect that individual independent variables have on cumulative cross-correlation results, we can say that it varies between the two models. Judging by the standardized coefficients, in case of the first model (CML_M1_X) PDT_AFR has the strongest influence, followed by driving performance described using

either AAVC_SWA or AAVC_LP and AAVC_VH. Similar ranking can be obtained in case of the second model with visual attention described by NG and AGD having the highest influence, followed by AAVC_SWA or AAVC_LP and AAVC_VH. We also have to note that AAVC_VH typically has the smallest influence and was also non-significant in two models: CML_M1_SWA in the city study and CML_M2_LP in the highway study. This small influence can be explained by the existing overlap between vehicle heading, steering wheel angle and lane position. However, as illustrated in Figure 4.19, situations exist when vehicle heading can complement steering wheel angle and lane position. Since vehicle heading often proved to be significant and no problems with multicollinearity have been observed, we decided to keep it in the models.

Table 4.24 summarizes the regression results for the per-glance model of cross-correlation results (PG_M_X).

model	road	standardized beta coefficients				R ²
		AGD	AAVC_LP	AAVC_SWA	AAVC_VH	
PG_M_SWA	highway	0.188		0.7164	0.1548	0.81
	city	0.1403		0.7805	0.1127	0.81
PG_M_LP	highway	0.1178	0.6613		0.1484	0.67
	city	0.0903	0.6921		0.1372	0.69

Table 4.24 Standardized beta coefficients for individual reference studies for model PG_M_X.

The first thing that we can notice from Table 4.24 is that this model on average explains less variation in the per-glance cross-correlation results (average R^2 including both steering wheel angle and lane position is 0.75) compared to the two models explaining the cumulative results (average $R^2 \geq 0.89$ for both models). This can be

explained using our discussion in the introduction (Section 1.1.2), where we state that not every glance will necessarily instigate decrements in driving performance. As a result, this may create a higher uncertainty (variability) in the per-glance results, which cannot entirely be accounted for with the proposed model. Nevertheless, we can see that the model still provides a fairly good explanation of the per-glance cross-correlation results, since it managed to explain a considerable portion of the variance: $\geq 67\%$ for lane position and 81% for steering wheel angle. We can see that the same trend observed with the cumulative results occurred here as well, with per-glance steering wheel angle cross-correlation results providing better fits than the per-glance lane position cross-correlation results.

If we look at the size of the effect that independent variables produce on per-glance cross-correlation results, we can see that driving performance variables have the strongest influence (AAVC_SWA or AAVC_LP). Conversely, the influences of average glance duration (AGD) and vehicle heading (AAVC_VH) change in their importance: for steering wheel angle AGD is more important, while for lane position AAVC_VH appears to be more important. However, overall their importance is similar.

4.3.6 Modeling the Effect of the Driving Environment

As we had a chance to see in Section 4.2 (pg. 200) the effect of the environment is present. The goal of this section is to model this effect on the cross-correlation results. Similar to the previous section, we accomplish this by creating regression models which include the type of the environment as another independent variable. The data from the two reference studies directly help in achieving this goal.

As before, two types of cross-correlation results are considered: cumulative and per-glance. Regarding cumulative results, two models can be created: one using PDT_AFR for describing visual attention and the other which uses AGD and NG. Please note that we will keep the same abbreviations for all the variables and models as in the previous section. We will also follow exactly the same procedure for conducting regression analyses as outlined in Section 4.3.5.

The first model for cumulative cross-correlation results (CML_M1_X) can be defined as follows:

$$XCORR_CML_SWA =$$

$$a_0 + a_1 \cdot PDT_AFR + a_2 \cdot AAVC_SWA + a_3 \cdot AAVC_VH + a_4 \cdot ROAD$$

$$XCORR_CML_LP =$$

$$b_0 + b_1 \cdot PDT_AFR + b_2 \cdot AAVC_LP + b_3 \cdot AAVC_VH + b_4 \cdot ROAD$$

Equation 4.4 Extending the cumulative CML_M1_X model to include driving environment.

The second cumulative model (CML_M2_X) can be defined as follows:

$$\begin{aligned}
 XCORR_CML_SWA &= \\
 &c_0 + c_1 \cdot AGD + c_2 \cdot NG + c_3 \cdot AAVC_SWA + c_4 \cdot AAVC_VH + c_5 \cdot ROAD \\
 XCORR_CML_LP &= \\
 &d_0 + d_1 \cdot AGD + d_2 \cdot NG + d_3 \cdot AAVC_LP + d_4 \cdot AAVC_VH + d_5 \cdot ROAD
 \end{aligned}$$

Equation 4.5 Extending the cumulative CML_M2_X model to include driving environment.

Finally, the regression model for per-glance cross-correlation results (PG_M_X) can be defined as follows:

$$\begin{aligned}
 XCORR_PG_SWA &= e_0 + e_1 \cdot AGD + e_2 \cdot AAVC_SWA + e_3 \cdot AAVC_VH + e_4 \cdot ROAD \\
 XCORR_PG_LP &= f_0 + f_1 \cdot AGD + f_2 \cdot AAVC_LP + f_3 \cdot AAVC_VH + f_4 \cdot ROAD
 \end{aligned}$$

Equation 4.6 Extending the per-glance PG_M_X model to include driving environment.

In all of the above models “X” represents either steering wheel angle (“SWA”) or lane position (“LP”). “ROAD” is a dummy independent variable which accounts for the driving environment and has two possible values: “city” and “highway.”

Regarding visual attention and driving performance variables, we expect to obtain significant and positive coefficients. However, regarding the ROAD variable the situation is somewhat different. Based on the cross-correlation results obtained in the two reference experiments, we had a chance to see an interesting effect: steering wheel angle cross-correlation results increased in city driving, while lane position cross-correlation

results decreased in city driving. In Section 4.2 (pg. 200) we provided an explanation which stated that the participants invested more effort on steering in the city environment, which resulted in less variation in the lane position. Conversely, less effort on steering in the highway environment resulted in larger variation in lane position. This suggests that the expected sign of the corresponding coefficient for the ROAD variable should be positive in case of steering wheel angle cross-correlation results (both cumulative and per-glance), while in case of lane position cross-correlation results the sign should be negative when the environment changes from highway to city.

Before starting the regression analyses, we had to transform all of our variables to improve the symmetries of their distributions. Since all the variables are positively skewed, we applied the same type of power transformation as in the previous section: X^q , where $q < 1$. Table 4.25 shows the variables proposed in our regression models (except ROAD) and the power exponents used in transforming those.

Variable name	Abbreviation	q (Cumulative, CML)	q (Per-glance, PG)
PDT Away From Road	PDT_AFR	0.4	-
Glance Duration	AGD	0.8	0.5
Number of Glances	NG	0.8	-
Steering Wheel Angle	AAVC_SWA	0.3	0.4
Lane Position	AAVC_LP	0.3	0.3
Vehicle Heading	AAVC_VH	0.3	0.3
Steering wheel angle cross-correlation	XCORR_(CML or PG)_SWA	0.4	0.5
Lane position cross-correlation	XCORR_(CML or PG)_LP	0.5	0.4

Table 4.25 Exponents of power transformations used for normalizing the data for pooled reference studies.

Figure 4.28 shows the distributions of both original (“O”) and transformed (“T”) independent and dependent variables which are used for creating regression models for per-glance cross-correlation results. We can see that power transformations improved the distributions of all variables.

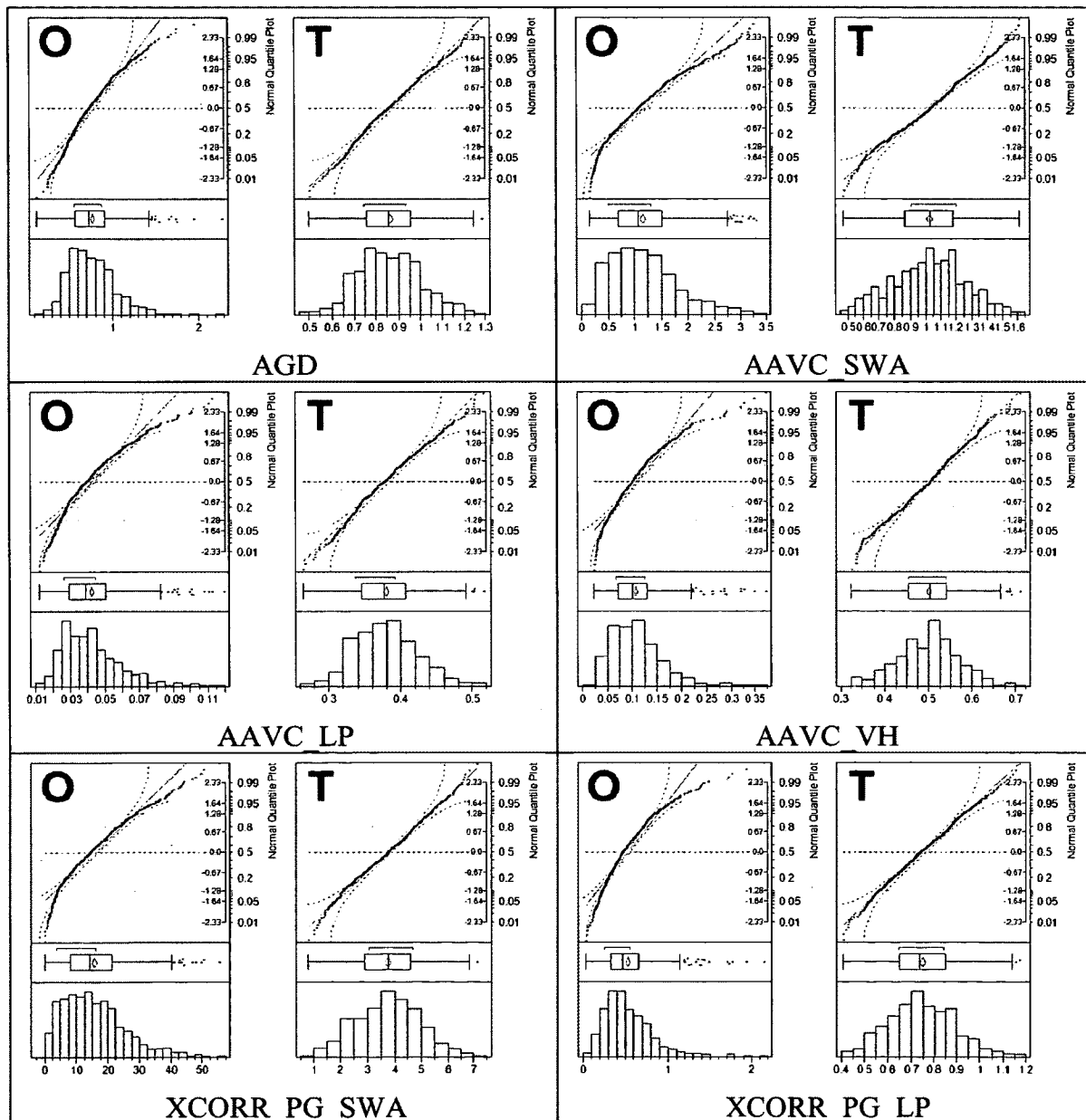


Figure 4.28 Distributions of dependent and independent variables used for modeling per-glance cross-correlation results for pooled highway and city studies.

Figures 4.29 and 4.30 depict the distributions of the independent and dependent variables used in modeling cumulative cross-correlation results, respectively.

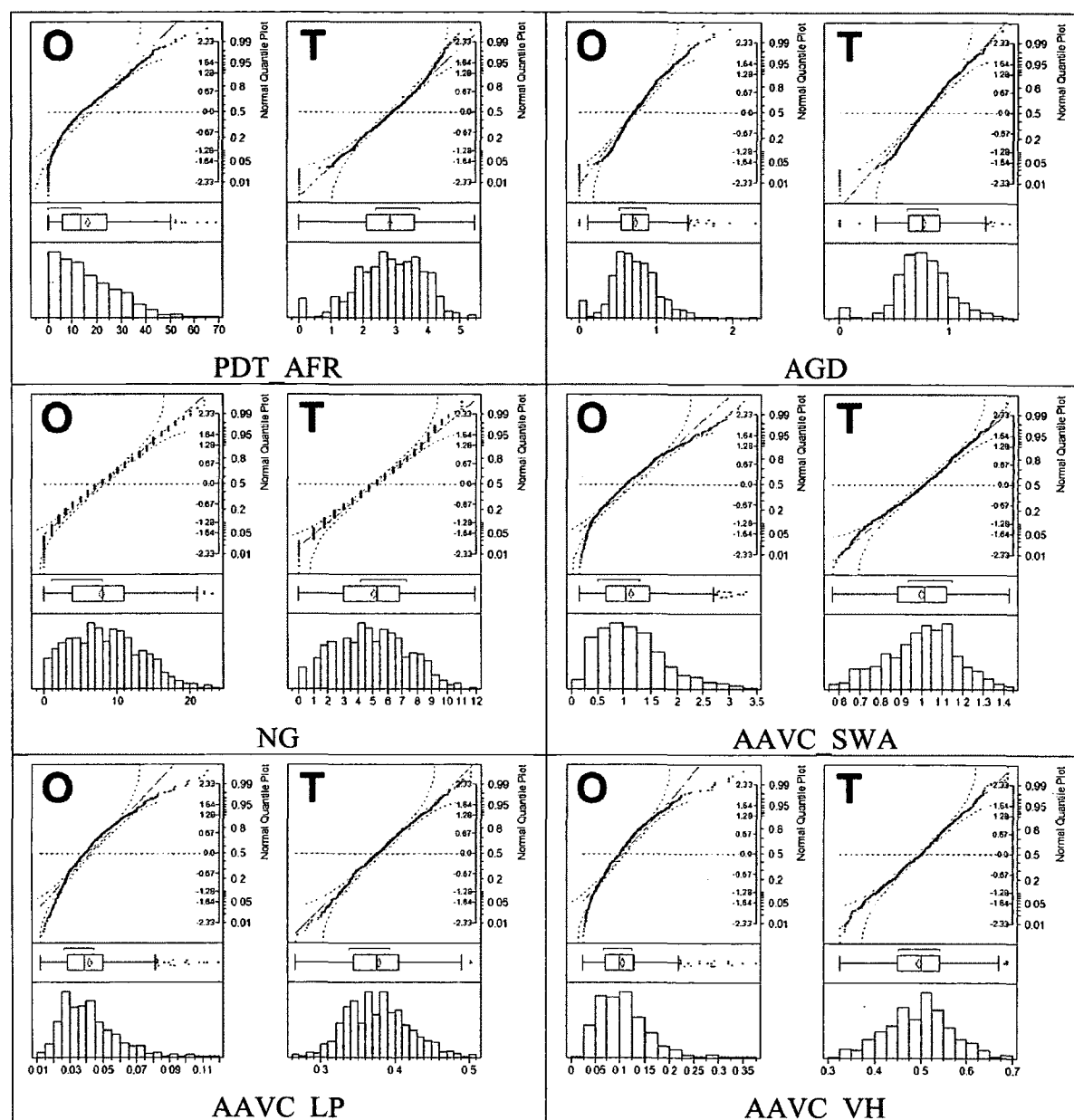


Figure 4.29 Distributions of independent variables used for modeling cumulative cross-correlation results for pooled highway and city studies.

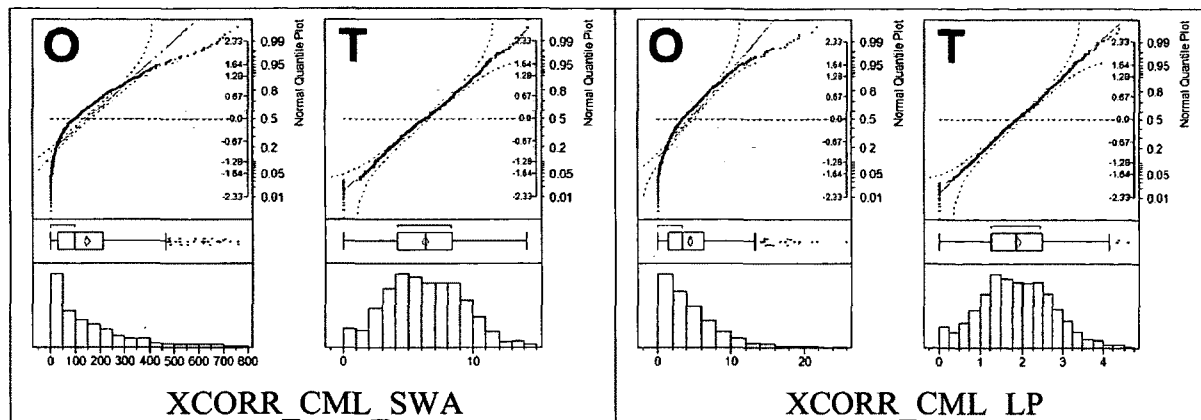


Figure 4.30 Distributions of dependent variables used for modeling cumulative cross-correlation results for pooled highway and city studies.

Modeling Cumulative Cross-Correlation Results

Cumulative steering wheel angle cross-correlation models: This section analyzes the following models for cumulative steering wheel angle cross-correlation results (CML_M1_SWA and CML_M2_SWA):

$$XCORR_CML_SWA =$$

$$a_0 + a_1 \cdot PDT_AFR + a_2 \cdot AAVC_SWA + a_3 \cdot AAVC_VH + a_4 \cdot ROAD$$

$$XCORR_CML_SWA =$$

$$c_0 + c_1 \cdot AGD + c_2 \cdot NG + c_3 \cdot AAVC_SWA + c_4 \cdot AAVC_VH + c_5 \cdot ROAD$$

Table 4.26 shows the regression results for the CML_M1_SWA model. As we can see, the model provided a very good fit, with approximately 91% of variance explained. All coefficient estimates are positive and significant, except AAVC_VH (p=0.7814).

CML_M1_SWA	R² = 0.91				
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	a_0	-7.3214	< 0.0001	-	-
PDT_AFR	a_1	1.7829	< 0.0001	0.616	1.21
AAVC_SWA	a_2	8.5816	< 0.0001	0.5053	1.3375
AAVC_VH	a_3	-0.2807	0.7814		
ROAD	a_4	0.168	< 0.0001	0.0574	1.1167

Table 4.26 Coefficient estimates for CML_M1_SWA model.

Table 4.27 shows the regression results for the CML_M2_SWA model. The model explained about 94% of variance and all coefficient estimates are positive and significant.

<i>CML_M2_SWA</i>	R² = 0.94				
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	c_0	-6.4038	< 0.0001	-	-
AGD	c_1	1.3584	< 0.0001	0.1147	1.4502
NG	c_2	0.7078	< 0.0001	0.5868	1.745
AAVC_SWA	c_3	7.0377	< 0.0001	0.4144	4.8009
AAVC_VH	c_4	2.0612	0.0195	0.0483	4.4421
ROAD	c_5	0.1194	0.0002	0.0408	1.2334

Table 4.27 Coefficient estimates for CML_M2_SWA model.

Cumulative lane position cross-correlation models: Two models for cumulative lane position cross-correlation results are analyzed here (CML_M1_LP and CML_M2_LP):

$$\begin{aligned}
 XCORR_CML_LP = & \\
 & b_0 + b_1 \cdot PDT_AFR + b_2 \cdot AAVC_LP + b_3 \cdot AAVC_VH + b_4 \cdot ROAD \\
 XCORR_CML_LP = & \\
 & d_0 + d_1 \cdot AGD + d_2 \cdot NG + d_3 \cdot AAVC_LP + d_4 \cdot AAVC_VH + d_5 \cdot ROAD
 \end{aligned}$$

Table 4.28 shows the results of the regression analysis for CML_M1_LP model. The model explained about 87% of variance. All coefficients regarding visual

attention and driving performance are significant and positive. Note that the coefficient for ROAD is also significant, but negative.

<i>CML_M1_LP</i>	R² = 0.87				
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	b_0	-2.8908	< 0.0001	-	-
PDT_AFR	b_1	0.587	< 0.0001	0.6918	1.299
AAVC_LP	b_2	5.5726	< 0.0001	0.2746	3.0192
AAVC_VH	b_3	2.0337	< 0.0001	0.1627	2.8385
ROAD	b_4	-0.0356	0.0237	-0.0414	1.7364

Table 4.28 Coefficient estimates for CML_M1_LP model.

Finally, Table 4.29 shows the regression results for the CML_M2_LP model. The model explained about 91% of variance. All coefficients are statistically significant and positive, except for ROAD which is negative.

<i>CML_M2_LP</i>	R² = 0.91				
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	d_0	-2.7606	< 0.0001	-	-
AGD	d_1	0.3198	< 0.0001	0.0921	1.4221
NG	d_2	0.2437	< 0.0001	0.6891	1.705
AAVC_LP	d_3	7.097	< 0.0001	0.3497	3.1587
AAVC_VH	d_4	0.9598	0.0002	0.0768	3.0342
ROAD	d_5	-0.0452	0.0006	-0.0527	1.7379

Table 4.29 Coefficient estimates for CML_M2_LP model.

Modeling Per-glance Cross-Correlation Results

Per-glance steering wheel angle cross-correlation model: We analyze the following model (PG_M_SWA) for per-glance steering wheel angle cross-correlation results here:

$$XCORR_PG_SWA = e_0 + e_1 \cdot AGD + e_2 \cdot AAVC_SWA + e_3 \cdot AAVC_VH + e_4 \cdot ROAD$$

The results of the regression analysis for the PG_M_SWA model are outlined in Table 4.30. The model explained about 83% of variance. All coefficients are positive and significant.

<i>PG_M_SWA</i>	R² = 0.83				
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	e_0	-2.3394	< 0.0001	-	-
AGD	e_1	1.2728	< 0.0001	0.1499	1.1091
AAVC_SWA	e_2	3.9317	< 0.0001	0.7664	4.6811
AAVC_VH	e_3	1.9499	0.0011	0.1132	4.4056
ROAD	e_4	0.0961	< 0.0001	0.0805	1.2641

Table 4.30 Coefficient estimates for PG_M_SWA model.

Per-glance lane position cross-correlation model: This section analyzes the following model for the per-glance lane position cross-correlation results (PG_M_LP):

$$XCORR_PG_LP = f_0 + f_1 \cdot AGD + f_2 \cdot AAVC_LP + f_3 \cdot AAVC_VH + f_4 \cdot ROAD$$

Table 4.31 presents the results of the regression analysis for the PG_M_LP model. The model explained about 72% of variance. The coefficients associated with the

driving performance and visual attention variables are positive and significant. The coefficient associated with the ROAD variable can be considered weakly significant at the $p=0.1$ level, so we decided to keep it in the model. We can also see that it has a negative sign, as was expected.

<i>PG_M_LP</i>	$R^2 = 0.72$				
Variable name	Coefficient	Estimate	p-value	Std. beta	VIF
Intercept	f_0	-0.3366	< 0.0001	-	-
AGD	f_1	0.0946	< 0.0001	0.0944	1.0905
AAVC_LP	f_2	2.3698	< 0.0001	0.7275	3.1114
AAVC_VH	f_3	0.2092	0.0023	0.103	2.5274
ROAD	f_4	-0.0064	0.1043	-0.0455	1.7512

Table 4.31 Coefficient estimates for PG_M_LP model.

Discussion of the Regression Results for Pooled Reference Studies

We started this section with the goal of modeling the effect of the driving environment on cross-correlation results. Table 4.32 (CML_M1_X model) and Table 4.33 (CML_M2_X model) provide an overview of the significant standardized coefficients for all variables used for modeling the cumulative steering wheel angle and lane position cross-correlation results. Similarly, Table 4.34 provides the same information, but for per-glance cross-correlation results.

model	standardized beta coefficients					R ²
	PDT_AFR	AAVC_SWA	AAVC_LP	AAVC_VH	ROAD	
CML_M1_SWA	0.616	0.5053		ns	0.0574	0.91
CML_M1_LP	0.6918		0.2746	0.1627	-0.0414	0.87

Table 4.32 Standardized beta coefficients for pooled reference studies for model

CML_M1_X.

model	standardized beta coefficients						R ²
	NG	AGD	AAVC_SWA	AAVC_LP	AAVC_VH	ROAD	
CML_M2_SWA	0.5868	0.1147	0.4144		0.0483	0.0408	0.94
CML_M2_LP	0.6891	0.0921		0.3497	0.0768	-0.053	0.91

Table 4.33 Standardized beta coefficients for pooled reference studies for model

CML_M2_X.

model	standardized beta coefficients					R ²
	AGD	AAVC_SWA	AAVC_LP	AAVC_VH	ROAD	
PG_M_SWA	0.1499	0.7664		0.1132	0.0805	0.83
PG_M_LP	0.0944		0.7275	0.103	-0.0455	0.72

Table 4.34 Standardized beta coefficients for pooled reference studies for model

PG_M_X.

Based on these results, the first important conclusion that we can draw is that the same set of proposed independent variables can also be used to describe the pooled cross-correlation results as when we analyzed the reference studies individually. The second conclusion provides another support for hypothesis H4: driving environment has a significant effect on both cumulative and per-glance cross-correlation results. Furthermore, we demonstrated that the change of the environment from highway to city driving increases steering wheel angle cross-correlation results, which is judged by the positive sign of the coefficient to the ROAD variable. Conversely, the same change in driving environment decreases lane position cross-correlation results, indicated by the negative sign of the coefficient to the ROAD variable. All other coefficients are positive, which indicates that the cross-correlation results increase as their corresponding variables increase. This conclusion agrees with the results that we obtained by analyzing each reference study individually.

Consistent with the individual regression analyses, we again obtained a somewhat better fit for the cumulative cross-correlation results when the second model was used: average R^2 for both steering wheel angle and lane position is 0.89 for the first model and 0.925 for the second model. Similarly, cumulative models provided better fits compared to the per-glance model for which the average R^2 is 0.775. The same explanations that we provided for these effects in the previous section can be applied here as well.

4.3.7 General Discussion and Future Direction

This chapter provided us with an important insight into the mechanisms which influence the behavior of our cross-correlation results. Namely, in agreement with our hypothesis H3, the regression analyses conducted in the previous sections revealed significant effects of the proposed variables. Specifically, we proposed two models for cumulative and one for per-glance cross-correlation results. The first cumulative model included PDT_AFR, AAVC_X (where “X” represents either “SWA” or “LP” based on whether the dependent variable is steering wheel angle or lane position cross-correlation result) and AAVC_VH variables. The second cumulative model included NG, AGD, AAVC_X and AAVC_VH variables. Finally, the per-glance model included AGD, AAVC_X and AAVC_VH. These models were applied individually for each reference study: highway and city driving. Drawing from the results obtained in the previous sections, we can make the following general conclusions:

1. The estimated coefficients that correspond to the proposed variables proved to be positive in all models, thus indicating that the increases in these variables directly contribute to the increases in cross-correlation results.
2. The same set of variables can be used for describing both steering wheel angle and lane position cross-correlation results (of course, appropriate driving performance variables should be used: a variable based on steering wheel angle should be used only with steering wheel angle cross-correlation results, for example). This conclusion suggests that the same underlying mechanisms influence changes in both types of cross-correlation results.

3. The same set of variables proved to be significant in both reference studies. This confirms the importance of the proposed variables, since they provided fairly good explanations of the cross-correlation results (judged by the coefficients of determination) in two unrelated experiments conducted under different experimental conditions.

These conclusions are important because they directly indicate how the in-vehicle interactions should be tuned in order to minimize (or at least reduce) the negative impacts on driving and cognitive load.

Proof-of-concept for the Predictive Ability of Cross-Correlation Results

Our regression models revealed a set of variables which significantly contribute to the cross-correlation results. It is important to note that our goal was not to obtain a universal model which could be applied in any arbitrary study and under any experimental conditions. Rather, our regression analyses were intended to facilitate the explanation and as the proof-of-concept for the predictive ability of the cross-correlation results. To demonstrate this concept, we applied regression models (cumulative, CML_M1_X and per-glance, PG_M_X) obtained in Section 4.3.6 based on our reference studies (“Highway Driving and iPod Interactions” and “City Driving and iPod Interactions”) to our navigation experiment presented in Chapter 3 (“Exploring Augmented Reality Navigation Aids”). We can recall that these models account for the effect of the driving environment by providing a separate “ROAD” variable. Since the navigation study was performed in the city environment, we set the “ROAD” variable to 1 (city=1, highway=0) in both models. We have to note here that the city environment employed in the navigation study was much more complex due to various confounding

variables (such as pedestrians, random ambient traffic and many consecutive turns at intersections) than the one used for obtaining these models. Nevertheless, they can still be used to demonstrate the general predictive ability. The following model (CML_M1_X) is applied for predicting the cumulative steering wheel angle and lane position cross-correlation results:

$$XCORR_CML_SWA = (-7.1534 + 1.7829 \cdot PDT_{AFR}^{0.4} + 8.5816 \cdot AAVC_SWA^{0.3})^{2.5}$$

$$XCORR_CML_LP =$$

$$(-2.9264 + 0.587 \cdot PDT_{AFR}^{0.4} + 5.5726 \cdot AAVC_LP^{0.3} + 2.0337 \cdot AAVC_VH^{0.3})^2$$

The per-glance cross-correlation results are predicted using the following model (PG_M_X):

$$XCORR_PG_SWA =$$

$$(-2.2433 + 1.2728 \cdot AGD^{0.5} + 3.9317 \cdot AAVC_{SWA}^{0.4} + 1.9499 \cdot AAVC_VH^{0.3})^2$$

$$XCORR_PG_LP =$$

$$(-0.343 + 0.0946 \cdot AGD^{0.5} + 2.3698 \cdot AAVC_LP^{0.3} + 0.2092 \cdot AAVC_VH^{0.3})^{2.5}$$

Please note that in order to make the predictions of the cross-correlation results using the above models, we have to account for the normalization transformations ($X' = X^q$) that we introduced to each of our independent and dependent variables (see Table 4.25). This explains the exponents that can be seen in the above equations.

Table 4.35 shows the observed cumulative cross-correlation results for each experimental condition in the navigation study (SV, AR, SPND). We show both the magnitudes of the most prominent peaks as well as the areas below the cross-correlation functions. We also calculated the ratios of the magnitudes between individual experimental conditions. By comparing the common ratios, we can see that they are very similar, thus indicating the equivalence of the results obtained using either peaks or areas.

Xcorr	Condition	Observed magnitudes of		Ratios of observed prominent peaks		Ratios of observed areas	
		prominent peaks	areas				
Cumulative SWA	SV	10.65	83.1667	SV/AR	3.97	SV/AR	3.88
	AR	2.682	21.4291	SV/SPND	1.41	SV/SPND	1.37
	SPND	7.53	60.6336	SPND/AR	2.81	SPND/AR	2.83
Cumulative LP	SV	0.1211	1.2788	SV/AR	3.45	SV/AR	3.37
	AR	0.0351	0.3791	SV/SPND	1.39	SV/SPND	1.37
	SPND	0.0873	0.9323	SPND/AR	2.48	SPND/AR	2.46

Table 4.35 Observed cumulative cross-correlation results in the navigation study.

Table 4.36 shows the results of the cumulative predictions obtained using the CML_M1_X model. Even though the predicted values do not match the observed values in absolute terms, what is very important to notice is that the ranking of the results is the same: SV obtained the highest score, followed by SPND and AR. We also calculated the ratios of the predicted results between all conditions. In all cases we obtained somewhat smaller results compared to the ratios of the observed values. Nevertheless, the ranking of the ratios for the predicted results matches the ranking of the ratios for the observed results.

Model	Condition	Predicted areas	Ratios of predicted areas	
CML_M1_SWA	SV	435.5752	SV/AR	1.7056
	AR	255.3766	SV/SPND	1.0915
	SPND	399.047	SPND/AR	1.5626
CML_M1_LP	SV	7.6813	SV/AR	1.8267
	AR	4.2051	SV/SPND	1.1043
	SPND	6.9556	SPND/AR	1.6541

Table 4.36 Predicted cumulative cross-correlation results using CML_M1_X model.

The fact that our cumulative model obtained the predictions which differ from the observed data is not surprising. There are two main reasons for this result: difference in the durations of experimental segments between the two studies and a much more challenging driving environment that the participants experienced in the navigation study.

If an experimental segment is long, there may be more opportunity for involvement in secondary tasks (of course, it is not guaranteed that the interaction will actually occur more often). This may result in larger cumulative cross-correlation results, since they provide an overall effect of secondary task engagements. Our model was created in the iPod study where the average segment duration was 39.44 seconds, while in the navigation study it was 7.36 seconds. Therefore, we decided to “normalize” our predicted cumulative results by the ratio of segment durations between the two studies: $39.44 / 7.36 = 5.36$. Table 4.37 gives the comparison of the observed and the “normalized” predicted results.

Model	Condition	Observed areas	Normalized predicted areas
CML_M1_SWA	SV	83.1667	81.264
	AR	21.4291	47.6449
	SPND	60.6336	74.4491
CML_M1_LP	SV	1.2788	1.4331
	AR	0.3791	0.7845
	SPND	0.9323	1.2977

Table 4.37 Comparison of observed and normalized predicted cumulative cross-correlation results.

We can see that the predicted values are now very close to the observed values. We can propose two reasonable explanations for the remaining differences. First comes from the unaccounted differences in the driving environment between the two studies, such as pedestrians and turns at intersections. And second, interactions with the navigation devices were visual-only as opposed to manual-visual with the iPod. Nevertheless, after the normalization our model provided a fairly good generalization for the navigation study.

Table 4.38 and Table 4.39 show the observed and predicted (using the PG_M_X model) per-glance cross-correlation results for the navigation study. If we compare the observed and the predicted results we can see that they are very close. This indicates that our per-glance model managed to generalize fairly well to an unrelated study. We have to recall here that the difference between SV and SPND was not significant when comparing per-glance steering wheel angle cross-correlation results. Very small differences in magnitudes of their predicted results reflect this finding. Furthermore, AR obtained the smallest predicted per-glance result, which agrees with the

observed result. Also, no differences have been detected between the three PNDs regarding per-glance lane position cross-correlation results, which can also be seen in the predicted results by noticing very small differences in the predicted magnitudes.

Xcorr	Condition	Observed magnitudes of		Ratios of observed prominent peaks		Ratios of observed areas	
		prominent peaks	areas				
Per-glance SWA	SV	7.022	53.2492	SV/AR	1.29	SV/AR	1.26
	AR	5.407	42.3158	SV/SPND	1.06	SV/SPND	1.03
	SPND	6.627	51.9176	SPND/AR	1.23	SPND/AR	1.23
Per-glance LP	SV	0.0749	0.7949	SV/AR	1.01	SV/AR	1.01
	AR	0.0739	0.7893	SV/SPND	0.99	SV/SPND	0.99
	SPND	0.076	0.8059	SPND/AR	1.03	SPND/AR	1.02

Table 4.38 Observed per-glance cross-correlation results in the navigation study.

Model	Condition	Predicted areas	Ratios of predicted areas	
PG_M_SWA	SV	53.6173	SV/AR	1.0477
	AR	51.1756	SV/SPND	0.9785
	SPND	54.7952	SPND/AR	1.0707
PG_M_LP	SV	0.9019	SV/AR	1.0455
	AR	0.8627	SV/SPND	0.9886
	SPND	0.9123	SPND/AR	1.0575

Table 4.39 Predicted per-glance cross-correlation results using PG_M_X model.

One may ask the following question here: why is it the case that the differences between the predicted and the observed results were much smaller when the per-glance model was used compared to the non-normalized cumulative model? The reason is in the fact that the per-glance result describes the effect of an *individual* instance of interaction.

Therefore, the per-glance result is inherently “normalized” and does not depend on the length of the experimental epoch and the frequency of involvement in the secondary task.

Comparing Predictions of Cross-Correlation Results between Reference Studies

We presented two reference studies which analyzed the same secondary task (interactions of various difficulties with an iPod) under different driving environments. As we had a chance to see in Section 4.2 driving environment produced significant effects on all measures. Therefore, it would be interesting to analyze how well our models generalize for the two reference studies. Specifically, we would like to predict the cross-correlation results from one reference study using a model derived from the other reference study.

In order to simplify the terminology, we will refer to the “Highway Driving and iPod Interactions” study as the “highway” study. Similarly, the “City Driving and iPod Interactions” study will be referred to as the “city” study.

For predicting the cumulative cross-correlation results in the highway study we used the following models derived from the city study:

$$XCORR_CML_SWA = (-5.2313 + 1.1432 \cdot PDT_AFR^{0.5} + 7.3513 \cdot AAVC_SWA^{0.4})^{2.5}$$

$$XCORR_CML_LP =$$

$$(-3.157 + 0.3479 \cdot PDT_AFR^{0.5} + 5.4499 \cdot AAVC_LP^{0.2} + 2.1844 \cdot AAVC_VH^{0.4})^2$$

For predicting the per-glance cross-correlation results in the highway study we used the following models derived from the city study:

$$XCORR_PG_SWA =$$

$$(-1.2638 + 1.0872 \cdot AGD^{0.6} + 3.2654 \cdot AAVC_SWA^{0.5} + 2.0054 \cdot AAVC_VH^{0.4})^2$$

$$XCORR_PG_LP =$$

$$(-0.9969 + 0.0802 \cdot AGD^{0.6} + 2.8738 \cdot AAVC_LP^{0.2} + 0.28 \cdot AAVC_VH^{0.4})^2$$

For predicting the cumulative cross-correlation results in the city study we used the following models derived from the highway study:

$$XCORR_CML_SWA =$$

$$(-4.605 + 0.8923 \cdot PDT_AFR^{0.4} + 4.3857 \cdot AAVC_SWA^{0.2} + 2.347 \cdot AAVC_VH^{0.2})^{3.3}$$

$$XCORR_CML_LP =$$

$$(-2.19 + 0.4644 \cdot PDT_AFR^{0.4} + 4.3136 \cdot AAVC_LP^{0.3} + 1.4266 \cdot AAVC_VH^{0.2})^{2.5}$$

Finally, for predicting the per-glance cross-correlation results in the city study we used the following models derived from the highway study:

$$XCORR_PG_SWA =$$

$$(-3.2091 + 2.0241 \cdot AGD^{0.2} + 2.9781 \cdot AAVC_SWA^{0.3} + 1.7171 \cdot AAVC_VH^{0.2})^{2.5}$$

$$XCORR_PG_LP =$$

$$(-1.0856 + 0.3481 \cdot AGD^{0.2} + 3.0039 \cdot AAVC_LP^{0.3} + 0.4516 \cdot AAVC_VH^{0.2})^{1.67}$$

Table 4.40 presents the observed and predicted cumulative and per-glance cross-correlation results for both reference studies using the above models.

Xcorr	Condition	Observed areas for highway study	Predicted areas for highway study using models derived from city study	Observed areas for city study	Predicted areas for city study using models derived from highway study
Cumulative SWA	B	189.02	207.06	331.38	258.99
	E	27.99	26.41	58.37	48.93
	D	108.87	113.31	154.32	110.05
Cumulative LP	B	8.36	7.12	6.54	7.14
	E	1.66	1.68	1.64	1.8
	D	4.54	4.12	3.57	3.39
Per-glance SWA	B	16.5	17.35	25.4	21.79
	E	7.59	8.47	11.89	11.14
	D	13.43	14.6	17.63	15.06
Per-glance LP	B	0.76	0.67	0.52	0.55
	E	0.47	0.46	0.34	0.39
	D	0.57	0.55	0.43	0.44

Table 4.40 Observed and predicted cross-correlation results for both reference studies.

Predicted results for each study were obtained using models from the opposite study.

Two conclusions can be derived based on the above results. First, models derived from both studies provided very good generalizations judging by the small differences between the predicted and the observed values. And second, the ranking of the experimental conditions (B, E, D) based on the predicted values always matches the ranking based on the observed values. High accuracy of the predicted results can be explained by the fact that both studies analyzed the same type of interaction. We

demonstrated in Section 4.2 that by changing the driving environment the distributions of the common variables (driving and visual attention measures) changed, however, our models managed to generalize very well outside of the scope of the data for which they were generated. This provides another source of evidence for the predictive ability of the cross-correlation results.

Future Direction

The results presented in this chapter strongly suggest that the outcomes of our cross-correlation method can be predicted using the proposed variables and are the first step towards obtaining a more general model. We believe that this goal can be achieved in the future studies by extending our “reference experiments” approach to include more driving and in-vehicle interaction conditions. Specifically, different road types (such as curvy and residential), time of day (such as daylight and night), weather conditions (such as rain, snow and fog) and the combinations of these should be investigated. This would provide additional data, which would allow fine tuning the models. Additionally, it may be of interest to also model the driving environment from the standpoint of events of interest, such as whether a pedestrian was crossing the street in front of the participant in each experimental segment, how many ambient vehicles were present in each segment, and so on. This information may provide additional insight into the environmental effect.

One reason why our current models may not be used directly in any arbitrary experiment is the fact that not all in-vehicle interactions are described well using visual attention. For example, conversing on a hands-free cell phone would require some other variable besides visual attention which has the potential of introducing changes in driving performance, such as the instants when the driver utters the first word. Similarly, gestures

[134] are becoming increasingly popular in the automotive environment. Visual attention would likely not work well in this case either, since gestures can be performed without directing visual attention away from the road. Therefore, more studies are necessary in order to obtain a complete understanding of how the cross-correlation results change under different driving environments and in-vehicle interactions.

Nevertheless, the fact that we obtained a set of variables which describe well the changes in the cross-correlation results coming from the particular manual-visual type of interaction with an iPod can help with design decisions. Specifically, we can argue that by minimizing those variables we can reduce the cross-correlation results. By making comparisons with subjective (NASA-TLX) and physiological (average skin conductance) estimates of cognitive load we demonstrated in the current and the previous chapter that the cumulative cross-correlation results are strongly related to cognitive load. Based on these results we can argue that the above minimizations would eventually contribute to decreases in cognitive load. The question is which of these variables can be practically minimized?

We had a chance to see that driving performance can partially be impacted by the driving environment. Demanding driving conditions produce higher cognitive load. However, the choice of the environment is often beyond a driver's decision. Therefore, the most obvious choice would be to reduce the amount of secondary task engagement. In case of manual-visual interactions this means reducing the amount of visual attention directed off-road, in other words, reducing PDT away from the road. As we had a chance to see, this can be accomplished either by reducing the total number of glances (NG) or the duration of individual glances (AGD). Reducing both variables simultaneously would

be the best option, because it is important to avoid cases where the drivers would decrease one variable and increase the other, which would essentially make the overall visual attention the same. This effect was actually observed in our reference experiments (see Section 4.2), where the participants directed a larger number of shorter glances off-road in the city driving and a smaller number of longer glances in the highway driving. This may lead to the usability paradox discussed in the background (Chapter 2): the drivers may feel that short individual interactions are safe and thus start performing more interactions.

Another way of reducing the cross-correlation results would be to minimize the changes in driving performance variables. Even though those changes are influenced both by the environment and the secondary task engagements, the technological advancements may allow us to decouple those influences. For example, advanced driver assistance systems in the forms of lane departure/keeping [135] and adaptive cruise control [136] are just some examples of the variety of products that are penetrating the automotive market. It may be possible in the near future to temporarily turn the control of the vehicle to an intelligent controller which would keep the vehicle steady in the center of the lane while the driver is attending to the secondary task. How well such a system may work can directly be tested using our cross-correlation method.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

As the new technological advancements are becoming available, an increasing number of in-vehicle services are being introduced in vehicles every day. Most of these services are meant to improve drivers' overall driving experience by enabling access to social networks, traffic reports and infotainment systems, to name just a few. There is plenty of evidence that drivers are finding value in using those services. Therefore, we can conclude that in-vehicle distractions are here to stay. Often equipment manufacturers try to reduce the negative effects of interactions with in-vehicle devices on driving by resorting to creative applications of various interaction modalities, such as spoken. Nevertheless, statistics indicate that driver distraction-induced crashes are on the rise [5]. This suggests that reliable tools are necessary for detecting the potential for distraction which would allow informing design decisions even before a device is introduced in vehicles.

The method presented in this dissertation offers one possible solution which can help in detecting and measuring distraction introduced by in-vehicle (secondary) tasks. Specifically, we set out to develop a new performance measure which provides more sensitivity to changes in cognitive load compared to standard average-based

measures. As was discussed in the introduction, it is often the case that changes in cognitive load are not obvious using the average-based driving performance measures, such as variances of lane position and steering wheel angle. This occurs despite the fact that changes may exist in visual attention and/or subjective estimates of cognitive load. As a result, this may create a wrong impression about the influence on driving and cognitive load of the analyzed in-vehicle interactions. Additionally, as Wickens suggests [10], it is necessary to obtain multiple sources of evidence pointing to the same conclusion in order to be able to avoid circular arguments.

The main problem with the average-based measures is that they are unable to use the information about the potential sources of the observed changes in driving performance. This may result in missing localized changes due to various factors: durations of the analyzed intervals, influences occurring infrequently or non-interaction related changes masking the relevant ones. However, just because the influence of an in-vehicle interaction is not detected in the averages, it does not mean that it is not present and that it should be neglected. It cannot be emphasized enough how important those localized changes can be especially with interactions that can be performed very often (such as interactions with an MP3 player). Those individual interactions may appear simple to the drivers and even encourage the engagement. However, as Lee and Strayer [45] suggest, this behavior may lead to a usability paradox, where each individual interaction appears simple, but more frequent engagement increases the overall risk.

To circumvent this problem, we proposed a novel method which is based on the mathematical function of cross-correlation. This method is initiator-based, which means that it accounts for the potential causes of changes in driving performance. This

way we are able to isolate the events of interest and analyze their impacts on driving and cognitive load. This is in contrast to the average-based measures, which pool all the data together thus running the risk of missing events of interest in the averages. The next section will provide an overview of the contributions provided by this dissertation.

5.1 Primary Contributions

There are three goals that we stated at the beginning of this research:

- (G1) Introduce a cumulative measure of a secondary task engagement on cognitive load,
- (G2) Introduce an instance-based measure of a secondary task engagement on cognitive load,
- (G3) Provide explanations for the mechanisms underlying the cumulative and instance-based measures.

Our goals have been stated with generalization being our ultimate aim. Therefore, we defined our proposed method such that it can be applied to various types of in-vehicle interactions, such as haptic, spoken, visual, and so on. However, as a first step towards developing a truly generalized method, we constrained this research to two interaction modalities which are the most commonly found in the automotive environment: visual-only and manual-visual. The following sections explain how each goal was addressed, while Section 5.3 discusses the ways of extending this research. Additionally, Section 5.4 proposes potential applications of our method, some of which can be used in non-automotive research areas as well.

5.1.1 Addressing Goal 1

Our first goal was to introduce a cumulative measure of a secondary task engagement on cognitive load. We hypothesized (H1) that this goal can be addressed through an initiator-based quantification of cumulative secondary task engagement. What this means is that our method uses a sequence of reference points (“initiator sequence”) which indicates the occurrence of secondary task engagements along with an appropriate performance measure (“performance sequence”) which can detect the effects of those engagements. The cumulative effects of those secondary tasks on performance (which is a measure of cognitive load) are then evaluated by applying the mathematical function of cross-correlation between the two sequences.

When analyzing visual and manual-visual interactions with in-vehicle devices, one proven effect that has a negative influence on driving is gazing away from the forward road. Therefore, in this case glances directed away from the road make an appropriate initiator sequence, while changes (absolute value of change (AVC) in our case) in steering wheel angle or lane position can be used as performance sequences. The main idea behind this is that while the driver is not looking at the road she is not aware of the situation in front of the vehicle; thus, a correction in the vehicle position may have to be applied once the gaze is returned to the road. This of course does not imply that every glance directed off-road necessarily results in decrements in driving performance. However, those decrements are more likely when the visual attention is not directed to the road. Since the cross-correlation function takes time into consideration even the localized influences, which may get unnoticed in the averages, are detected.

Based on the results of three driving simulator studies (“Exploring Augmented Reality Navigation Aids” which was published at MobileHCI 2011 conference [36], “Highway Driving and iPod Interactions” and “City Driving and iPod Interactions”) presented in Chapters 3 and 4 we can conclude that H1 is supported. The overall effects of looking away from the road were revealed through prominent, statistically significant cross-correlation peaks for both driving performance sequences (AVC of steering wheel angle and lane position) and on average appeared about 0.6 seconds after the gaze returned to the road. These results clearly indicate that the effects of interactions with the tested in-vehicle devices (in our case personal navigation devices and iPod) are followed by pronounced changes in driving performance. Furthermore, by applying the two methods proposed in hypothesis H_{RP} (magnitudes of most prominent peaks and areas below the curves) we revealed statistically significant differences in cross-correlation results between different experimental conditions. This provided support for hypothesis H_{RP} , allowed ranking of the experimental conditions and also demonstrated high sensitivity that our method provides, given that the differences were not always detected using the average-based measures. For example, in the navigation study (“Exploring Augmented Reality Navigation Aids”) none of the average-based driving performance measures detected a significant effect of the navigation type. Similarly, in the iPod study which involved city driving (“City Driving and iPod Interactions”) no effect has been detected on the variance of lane position.

If we take into account the nature of the driving performance measures that we are using in our method (AVC of steering wheel angle and lane position), the cumulative cross-correlation result represents the amount of cumulative angular (for steering wheel

angle) and positional (for lane position) change introduced over the course of interaction with the secondary task. In other words, the cumulative result describes the overall effect of the secondary task engagement, which is exactly what we intended to accomplish in our first goal. In that respect our cumulative results are similar in nature to the “standard” measures which are known to be able to reflect changes in cognitive load, specifically, subjective, physiological and average-based measures. Therefore, it is beneficial to compare the ranking obtained using our method with the rankings obtained from these standard measures. If the rankings are the same, it confirms that both types of measures provide conclusions in the same direction, which provides support for the construct validity of our method.

We analyzed how cumulative cross-correlation results compare to subjective estimates of cognitive load using linear regressions. In all studies we obtained strong relationships ($R^2 > 0.88$), which indicate that both cumulative cross-correlation and NASA-TLX results point to the same conclusion regarding cognitive load changes. Table 5.1 gives an overview of the coefficients of determination in each study for both cumulative steering wheel angle (R_{sw}) and lane position (R_{lp}) cross-correlation results.

Study	R^2 for NASA-TLX vs. R_{sw}	R^2 for NASA-TLX vs. R_{lp}
Exploring Augmented Reality Navigation Aids	0.9749	0.9752
Highway Driving and iPod Interactions	0.9821	0.9608
City Driving and iPod Interactions	0.8818	0.9325

Table 5.1 Coefficients of determination obtained between NASA-TLX and cumulative cross-correlation results for three studies.

As we discussed in Chapter 2, other estimates of cognitive load exist, such as physiological measures. It is for that reason that we decided to compare our cumulative results to average values of heart rate and skin conductance. Both of these physiological measures were collected in the study which analyzed iPod interactions in city driving. Average skin conductance provided more sensitivity to changes in interaction type compared to heart rate. Therefore, we conducted a linear regression analysis between our cumulative cross-correlation results and average skin conductance. As we can see in Table 5.2, we again obtained very strong relationships ($R^2 > 0.81$).

Study	R^2 for average skin conductance vs. R_{sw}	R^2 for average skin conductance vs. R_{lp}
City Driving and iPod Interactions	0.8108	0.8747

Table 5.2 Coefficients of determination obtained between average skin conductance and cumulative cross-correlation results.

Finally, we can also compare our cumulative cross-correlation results to average-based measures of driving performance. If we recall from Section 4.2.4, among all average-based measures, variance of steering wheel angle provided the highest sensitivity towards changes in interaction type with the iPod (B, E, D) in both reference studies. It detected differences between all interaction types and the obtained ranking matched the one obtained using our cumulative cross-correlation results. This provides another source of support that construct validity is supported.

It is also worth mentioning that the variance of lane position was not very sensitive and detected only one difference in the highway driving study (between E and D conditions) and did not even detect the main effect of the interaction type in the city

driving study. Conversely, our cumulative lane position cross-correlation results detected significant differences between all interaction conditions in both reference studies.

Based on all of the above results we can conclude that goal G1 is accomplished. We had a chance to see that our method does provide a cumulative effect of secondary task engagements (H1), allows ranking of the experimental conditions based on these results (H_{RP}), has the potential to provide higher sensitivity compared to average-based measures, and also satisfies construct validity.

5.1.2 Addressing Goal 2

Our second goal (G2) was to introduce an instance-based measure of a secondary task engagement on cognitive load. We hypothesized (H2) that this goal can be addressed through initiator-based quantification of instances of secondary task engagement. In this case the same assumptions and data sequences were used as in H1. However, a normalization procedure was introduced in order to estimate the effects of individual instances of engagement. As already discussed, in our case instances of engagement included glances directed away from the forward road. Therefore, by addressing the second goal we obtained the average effects on driving and cognitive load introduced by individual glances.

We tested this hypothesis using the same driving simulator studies employed for testing H1 (Chapters 3 and 4). The effects of individual glances directed away from the road were revealed through prominent, statistically significant cross-correlation peaks. Similar to the cumulative results, the peaks on average appeared about 0.6 seconds after the gaze returned to the road. These results indicate that deteriorations in driving

performance exist even at the level of individual instances of engagement in the secondary task. This is a very important finding, because it agrees with our discussion that even those individual, local influences should not be neglected, since they have the potential to impact driving and cognitive load. Based on these results we can conclude that H2 is supported.

Using the two methods proposed in hypothesis H_{RP} (magnitudes of most prominent peaks and areas below the curves) we also demonstrated that instance-based results allow ranking. For example, in the navigation study, the instance-based (per-glance) steering wheel angle cross-correlation results revealed that SV and SPND produced similar effects when looked at from the standpoint of individual instances of engagement. However, AR PND produced a significantly smaller effect compared to both SV and SPND. Similarly, in case of iPod interaction (reference) studies we revealed large differences between all three interaction types (B, E, D) at the level of individual instances of engagement. These differences at the elementary levels of interaction are very important, because they allow us to learn something about the nature of the influence of individual glances that eventually give rise to the observed cumulative effects. Such low-level insight into different interaction types is impossible to obtain using the average-based measures, since they consider the experiment as a whole.

The above results demonstrate that we successfully quantified the effects of individual instances of engagement in the secondary tasks (H2) and also managed to rank those effects (H_{RP}) for different interaction types. This indicates that G2 is accomplished.

5.1.3 Addressing Goal 3

Finally, our last goal was to propose explanations for the mechanisms underlying the cumulative and instance-based measures. This goal was addressed by our third hypothesis (H3) which proposed that the following variables may have an important influence on our cross-correlation results: percent dwell time spent looking away from the road (PDT_AFR), average glance duration (AGD), number of glances (GN) and the average absolute amount of change in lane position (AAVC_LP), steering wheel angle (AAVC_SWA) and vehicle heading (AAVC_VH). Since our cross-correlation results represent a unique combination of both driving performance and visual attention, we selected the above variables in an attempt to provide the best descriptions of the both worlds. In order to analyze the effects of these variables we conducted a series of detailed multivariate regression analyses presented in Chapter 4. Given that cumulative and instance-based measures address different aspects of secondary task engagement, we created a separate model for each measure. The cumulative model included PDT_AFR (or alternatively AGD + GN), AAVC_X (where “X” represents either steering wheel angle (SWA) or lane position (LP)) and AAVC_VH. The instance-based model included AGD, AAVC_X and AAVC_VH.

We tested these proposed models using the data obtained from the two iPod interaction studies (“Highway Driving and iPod Interactions” and “City Driving and iPod Interactions”). These studies were designed using a “reference studies” approach, which means that they were fairly well controlled in order to minimize confounding variables. They included the same secondary task (three levels of interactions with an iPod) and the

only major difference consisted in the employed driving environment: highway in the one case and city in the other.

The results indicated statistically significant effects of all of the above variables on cross-correlation results. Specifically, all variables contributed with positive signs, thus indicating that the cross-correlation results increase as the values of these variables increase. The models provided very good fits to the data, with coefficients of determination ranging from 0.85 to 0.94 for cumulative models and from 0.69 to 0.81 for instance-based models. These are important results, because they indicate that these variables should be taken into consideration when designing in-vehicle devices. Specifically, the designers should strive to reduce these variables in order to reduce the impacts on driving and cognitive load. Based on these results we can conclude that G3 is accomplished as well.

We have to note here that our main intention in addressing G3 was to propose the underlying mechanisms which influence our cross-correlation results and not to create a comprehensive model which would address every possible in-vehicle interaction. However, we wanted to observe how the model obtained using our reference studies would predict the results obtained in an unrelated study, specifically, the navigation study presented in Chapter 3 (“Exploring Augmented Reality Navigation Aids”).

As we had a chance to see in Section 4.3.7, even though the predicted ranking of the experimental conditions was correct, our model predicted much higher values for the cumulative results compared to the observed results. We asserted that this was caused by the fact that our model was estimated for the reference studies which had much longer segment durations compared to the navigation study. After applying the normalization

factor (equal to the ratio of segment durations between the navigation and the reference studies) our predicted cumulative results matched the observed cumulative results very closely.

Regarding the instance-based results our model provided predictions which were fairly close to the observed results. The reason why no normalization was necessary in this case is the fact that our instance-based measure describes the effects at the level of individual instances of interaction.

Furthermore, we also compared how the individual models derived from the two reference studies would predict each other's results. In other words, we used a model derived from one study and used it to predict the results of the other study. As we saw in Section 4.3.7 (Table 4.40) both models provided predictions of the cumulative and instance-based results which very closely resembled the observed results.

The above results indicate that our models provide fairly good generalizations when applied to data from unrelated studies. This is important, since it indicates that the outcomes of our method are indeed predictable. Nevertheless, further studies are necessary to fine tune the models for more diverse driving and in-vehicle interaction conditions (Section 5.3 proposes more ideas in this direction). In any case, we can conclude that our results successfully demonstrated a proof-of-concept for the predictive ability and are the first step toward obtaining a more general model.

5.2 Secondary Contributions

Besides the main contributions outlined in the previous section, there are multiple secondary contributions that have been obtained in the studies presented in this dissertation that are worth mentioning:

1. At the end of Chapter 3 we hypothesized (H4) that driving environment may have a significant effect on driving and cognitive load. Our “reference study” approach allowed us to confirm H4 by observing a significant effect of the driving environment on all collected measures: visual attention, average-based and cross-correlation results. This suggests that the driving environment is an important factor and should be taken into account as well when analyzing in-vehicle interactions.
2. Our navigation study (“Exploring Augmented Reality Navigation Aids”, Chapter 3), published at MobileHCI 2011 conference [36], explored two novel PNDs: augmented reality and street view. Our results indicated that augmented reality provided for better visual attention, driving performance, and subjective preference compared to street view and standard map-based PNDs. Based on these results we can say that AR PND stands out as a safe and agreeable PND. Nevertheless, our participants brought to our attention two concerns that merit further study. First, our implementation of an AR PND did not provide global navigation information; it only informs drivers about the current route to follow. Three (of 18) participants in fact indicated they would have appreciated receiving information about upcoming turns. One approach to address this issue is that proposed by Kim and Dey [28]. Second, overlaying routes for long stretches of road may be distracting.

Two participants stated that they disliked the semi-transparent navigation route in the AR PND because it was always present in their peripheral vision. Showing AR directions only when a turn is coming may alleviate this problem.

3. Even though iPod interactions have been analyzed in the research literature [18;26], our reference studies provide new insight into those interactions. It is often the case that complex interactions require more steps in order to obtain a desired result. In that respect it is expected that complex interactions would produce larger negative overall effects on driving and cognitive load. However, according to our instance-based measure we now know that even the effects of individual instances of interaction differ based on the difficulty of the performed task. Namely, glances directed off-road are “elementary” units of interaction which are common for each interaction type. However, our results indicate that even at the level of individual glances the effects of different interaction types are not the same. This result is very important and cannot be obtained using average-based measures.

5.3 Extensions of the Current Work

This section presents ideas about further extensions of our proposed method.

The way our method is defined (initiator-based approach) lends itself well to extensions to other types of in-vehicle interactions. We discussed that this research was constrained to interactions which depend on visual and manual-visual modalities. In this case the proper initiator sequence is visual attention. However, for other modalities, such as spoken, this may not be the case. Conversing on a hands-free cell phone is one obvious

example which does not depend on visual attention. In such a case one possible initiator sequence may contain reference points whenever the driver first starts talking. Of course, other alternatives are possible as well depending on the event of interest that should be analyzed. In a similar fashion we can envision analyzing other interaction types, such as tactile where a driver can press buttons or produce gestures on touch-sensitive surfaces without removing eyes from the road.

Since we analyzed visual and manual-visual interactions, we used glances directed off-road for generating the initiator sequence. It may be interesting to explore whether the results would change if instead of glances we would use fixations. This is a reasonable question to ask, since fixations are limited in both temporal and spatial domain, while glances are concerned with general observations of the objects of interest.

In our approach we transformed steering wheel angle and lane position using the absolute value of change function in order to describe changes in driving performance. In doing so we did not distinguish between moving to the left or right: we assumed equal cost of colliding with parked or vehicles from the opposite direction. However, it may be possible to decouple the two sides and apply different weights depending on the direction of the turn. The assumption behind this is that the drivers may apply a different amount of correction based on the direction of the turn.

One aspect of our results that should be investigated further is the behavior of the lag of the most prominent peaks. As we hypothesized in Section 3.2.3, the lag of the most prominent peak may be related to the urgency of the situation in front of the vehicle and the reaction time. However, this assertion should be investigated in more experimental studies. It is possible that the introduction of more diverse experimental

conditions will provide more variability in lags, which may help in casting more light on this issue. In our scenarios the participants were not rushed to perform the correction. Also, whenever they engaged in secondary tasks they always had the road in their peripheral vision, which likely helped with keeping the vehicle on the road. However, it would be interesting to analyze whether the lag would change under more urgent conditions. For instance, a more urgent situation would be created if a driver would be forced to look way down on the central console in order to perform a task. This would completely eliminate the road from the periphery, thus potentially requiring a faster reaction when the gaze is returned back to the road. This situation is not very difficult to imagine, since HDDs are sometimes mounted fairly low or drivers may sometimes place their PNDs next to the gear shifter. Another way to test the effect of urgency would be to use the visual occlusion paradigm [137], where the picture of the road is switched off for predefined periods of time. The moments when the picture of the road appears can then be used as the initiator sequence for the cross-correlation analysis. By changing the duration of the occlusion and the type of the road (for example curvy) we may be able to change the urgency and see how it influences the lags and the peaks as well.

We demonstrated that our cross-correlation results closely follow the subjective estimates of cognitive load obtained through NASA-TLX questionnaires. Furthermore, we demonstrated a strong relationship with one physiological estimate, specifically, average skin conductance. As part of applying our method to other types of in-vehicle interactions and driving environments, it would be of interest to further analyze the relationship between our results and the above estimates of cognitive load in order to determine the exact shape of the relationship (i.e. linear or non-linear).

Finally, it may be worth exploring the applicability of our method to physiological measures. Namely, since our method has been defined in a way that it can be applied between any two sequences, it would be possible to cross-correlate changes in physiological measures with a desired initiator sequence. The advantage of this approach, compared to standard average-based measures, would be in revealing how the physiology changes over time and where the largest change is focused (peak and lag observed with the cross-correlation functions). Given that the physiological measures are also commonly analyzed using the average-based measures, the possibility of missing localized changes exists here as well.

5.4 Applicability of the Current Work

Given its generalized definition, there are many potential areas which may benefit from using our method. Some of those are outlined in this section.

Exploring Cognitive Load in Human Dialogues

In-vehicle spoken dialogue systems are gaining in popularity. However, it is not always clear which system behaviors might result in increased driver cognitive load, which in turn could have negative safety consequences. We conducted a preliminary study [138] to explore the use of pupil diameter coupled with our cross-correlation method in the evaluation of the effects of different dialogue behaviors on the cognitive load of the driver.

It has been shown in the literature that pupil diameter can be sensitive to changes in cognitive load [139]. Given pupil diameter's fast dynamics we expected that it would be sensitive to rapidly changing behaviors occurring during a dialog. For this

study, we used a less-structured task, which we felt is more representative of future HMI interaction than highly structured tasks (e.g. question-answer tasks). Specifically, we examined whether we can detect differences in cognitive load between times when the driver is engaged in a verbal game with a remote conversant, and after the game finishes. Our hypothesis was that, once a game finishes, drivers would experience reduced cognitive load, and that this would be reflected in decreased pupil diameter.

Pairs of participants (the *driver* and the *other conversant*) were engaged in a spoken dialog. The driver operated the driving simulator, while the other conversant was seated in another room. The participants communicated over the headphones. A total of six pairs participated in the experiment.

The spoken task was the game of “Taboo,” a game in which the other conversant is given a word, and needs to work with the driver to identify it, but cannot say that word or five related words. Participants played a series of Taboo games. We provided the words to the other conversant by displaying them on an LCD monitor. We imposed a time limit of 1 minute on each game. The experimenter signaled the end of each Taboo game with an audible beep (0.5 second long, high pitched sine wave) heard by both conversants. The end of a game was reached in one of three ways: when the driver correctly guessed the word, when the other conversant used a taboo word, or when the conversants ran out of time.

Figure 5.1 demonstrates the experimental setup. Even though the equipment allowed the participants to see each other, this condition was not analyzed in this preliminary study. Rather, we focused on voice-only interactions.

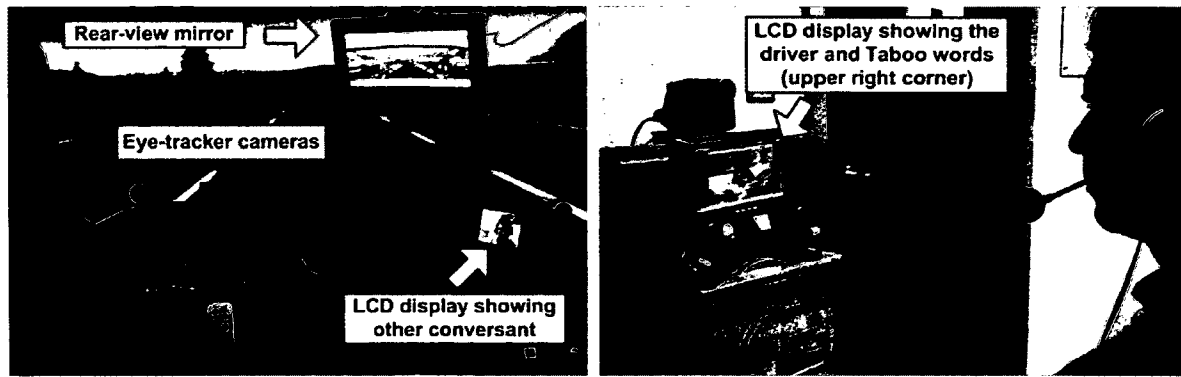


Figure 5.1 Driver (left) and other conversant (right).

Using the time instants when the beeps started, we segmented each experiment into individual games. We performed calculations and analyzed changes in cognitive load based on the pupil diameter data for each individual game. We estimated the cross-correlation function between the beep sequence (BS) and the pupil diameter sequence (PDS). BS is a sequence of 0s and 1s, where a '1' represents the moment when the beep started, signaling the end of a Taboo game. The PDS represents the processed measurements of the driver's left eye pupil diameter. We processed the raw measurements by interpolating short regions where the eye-tracker did not report pupil diameter measures, as well as by custom nonlinear smoothing (e.g. to reduce erroneous dips in pupil diameter caused by blinks).

The left graph in Figure 5.2 shows the average cross-correlation function for all subjects between the BS and the PDS. As hypothesized, the cross-correlation function drops in the seconds after the beep is initiated (which is at lag = 0 in the figure). The fact that the cross-correlation drops for about 5 seconds is consistent with the fact that the first contribution by the other conversant started on average about 4.6 seconds after the beginning of the beep (at lag = 0). The cross-correlation functions of two of the six drivers in this study did not clearly support our hypothesis. A number of factors could be

responsible, including differences in how the game was played by these participants (e.g. how engaged they were), and the noisiness of the pupil diameter measurements.

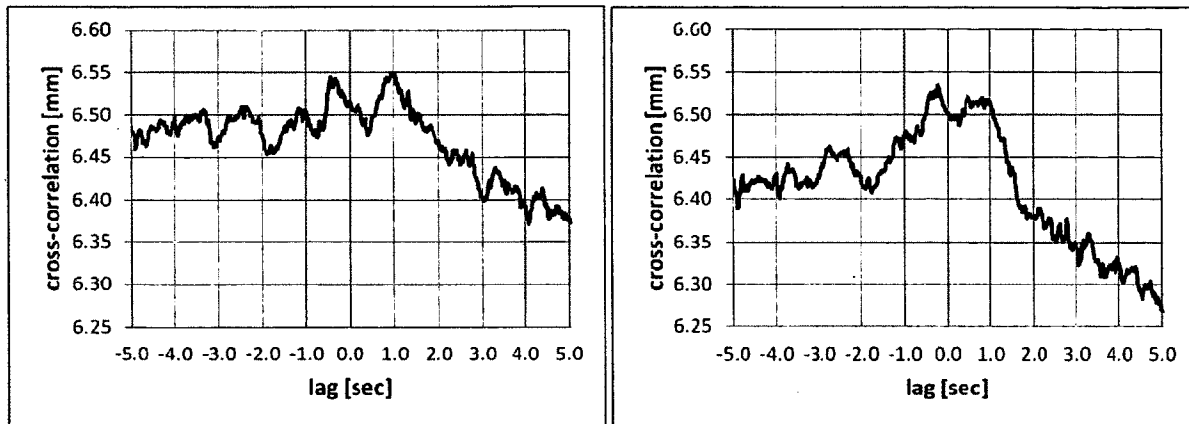


Figure 5.2 Cross-correlation functions for all six drivers (left) and for the four drivers whose results clearly supported our hypothesis (right).

Figure 5.2 (right) shows the average cross-correlation function for the four drivers whose data did in fact support our hypothesis. In comparison to the right graph, we can see that the drop is even more prominent. Additionally, the pupil diameter appears to be rising in the seconds before the end-of-game beep. We hypothesize that this rise is related to increased cognitive activity by the driver as she is attempting to find the word described by the other conversant. As correctly identifying this word is the most common cause of the end of the game, and thus the beep, it appears likely that cognitive load would indeed peak before the beep, thus at a negative value of lag. We should also expect to see a peak each time the driver makes a guess, but those peaks are not aligned with each other in time. Thus, they would not be visible after the cross-correlation operation.

The results of this preliminary study support our hypothesis that pupil diameter can be used to identify major changes in cognitive load during a dialogue. Furthermore,

this study demonstrates the applicability of our cross-correlation method to measures outside of the automotive domain.

Analyzing In-Vehicle Warnings

Various advanced driver assistance systems (ADAS) are either already present or are being introduced in vehicles nowadays [140], such as lane departure, blind spot and driver alert warning systems. These systems typically produce either an audible or a visual warning signal which indicates an imminent danger. It would be interesting to use the instants when those warning signals are issued as an initiator sequence in our cross-correlation analysis and analyze the changes in driving performance that may occur as a result. We would expect that two types of changes may be observed: intentional (if the danger is really obvious) or non-intentional (if the driver is confused about what is causing the danger and trying to determine where the danger is coming from). Similarly, our cross-correlation method would enable comparisons of different implementations of the same warning system with respect to the effect of the warning on driving (magnitudes of the most prominent peaks) and the urgency of the reaction to the warning (lags of the most prominent peaks).

Conversing on Hands-Free Cell Phone

It is well known that talking on both hand-held and hands-free cell phones negatively influences driving performance [19]. However, the average-based approach that the researchers typically employ characterizes the influences of those interactions on driving performance from the high level perspective. In other words, this way we can observe only the overall impact of the dialogue on driving. Using our cross-correlation method it would be possible to isolate the specific parts of the conversation which

contribute the most to decrements in driving performance. Similarly, we could isolate just the effects of dialing the phone and compare those to the effects of the conversation itself.

Exploring the Influences of Out-of-the-Vehicle Distractions

One area which appears under-researched [56] is the influence of out-of-the-vehicle distractions on driving performance. Some typical examples of external distractions are advertising, signs and even automobile accidents. Our cross-correlation method is well suited to extract the effects of these distractions by observing the instants when a driver glances towards those.

Analyzing Speech User Interfaces

There exists ample evidence in the literature that speech may be the preferred choice of interaction in vehicles [100]. However, the automatic speech recognition (ASR) engines are still not perfect, which prevents using ambient recognition. Rather, press-to-talk (PTT) buttons have to be used still, at least for indicating the beginning of an utterance. It would be useful to test the effects of using this button through our method. Specifically, it would be of interest to observe how the effects on driving change depending on the location of the PTT button. Tests could include fixed and location-free PTT buttons (such as the custom glove with an embedded PTT button used in [141]), which can be operated on curvy roads (require rotating the steering wheel) and on straight roads (do not require rotating the steering wheel). The results obtained from these studies would help in proposing or testing design choices.

Fatigue Effects on Driving Performance

It has been shown in the literature that fatigue negatively influences driving [142]. One of the easiest and most useful ways of detecting fatigue is through eyelid

closure. Wierwille et al. [143] derived a measure called PERCLOS, which reports the proportion of time per minute that the driver's eyes are at least 80% closed. In addition to that information, it would be useful to observe how large the impacts of individual eye-closures on driving performance are. This can also be accomplished using our cross-correlation method.

APPENDIX A

DATA SYNCHRONIZATION

This chapter provides an overview of a software application and hardware equipment which I designed for the purpose of data synchronization. Figure A.3 shows the equipment which is typically used in our experiments. Since all equipment maintains individual data collections a solution was needed which would enable seamless synchronization between all available data collections.

The main component in Figure A.3 is the driving simulator's control computer (a.k.a. HyperDrive) which is connected to the simulator through a local area network (LAN). This computer is used for creating scenarios, starting/stopping simulations and retrieving data from the simulator after concluding the experiment. Other equipment typically includes eye-tracker(s) (FaceLab 5.0 by SeeingMachines) and a physiological measurements monitor (ProComp Infinity by Thought Technology).

In an early attempt at synchronizing the simulator and the eye-tracker, we used a third-party application called NTP Fast Track [144], which synchronizes internal clocks of the computers of interest over the network. However, this solution proved to be unsatisfactory. Namely, if there is a need to restart or turn some of the computers off, the

clocks would fall out of sync fairly quickly. Conversely, it takes a considerable amount of time (ranging from minutes to hours) to get the computers back in sync, because NTP Fast Track adjusts the clocks by applying very small offsets over a long period of time. Thus, we needed another solution which would not depend on the computers' internal clocks. Furthermore, our physiological monitor cannot be synchronized over the network.

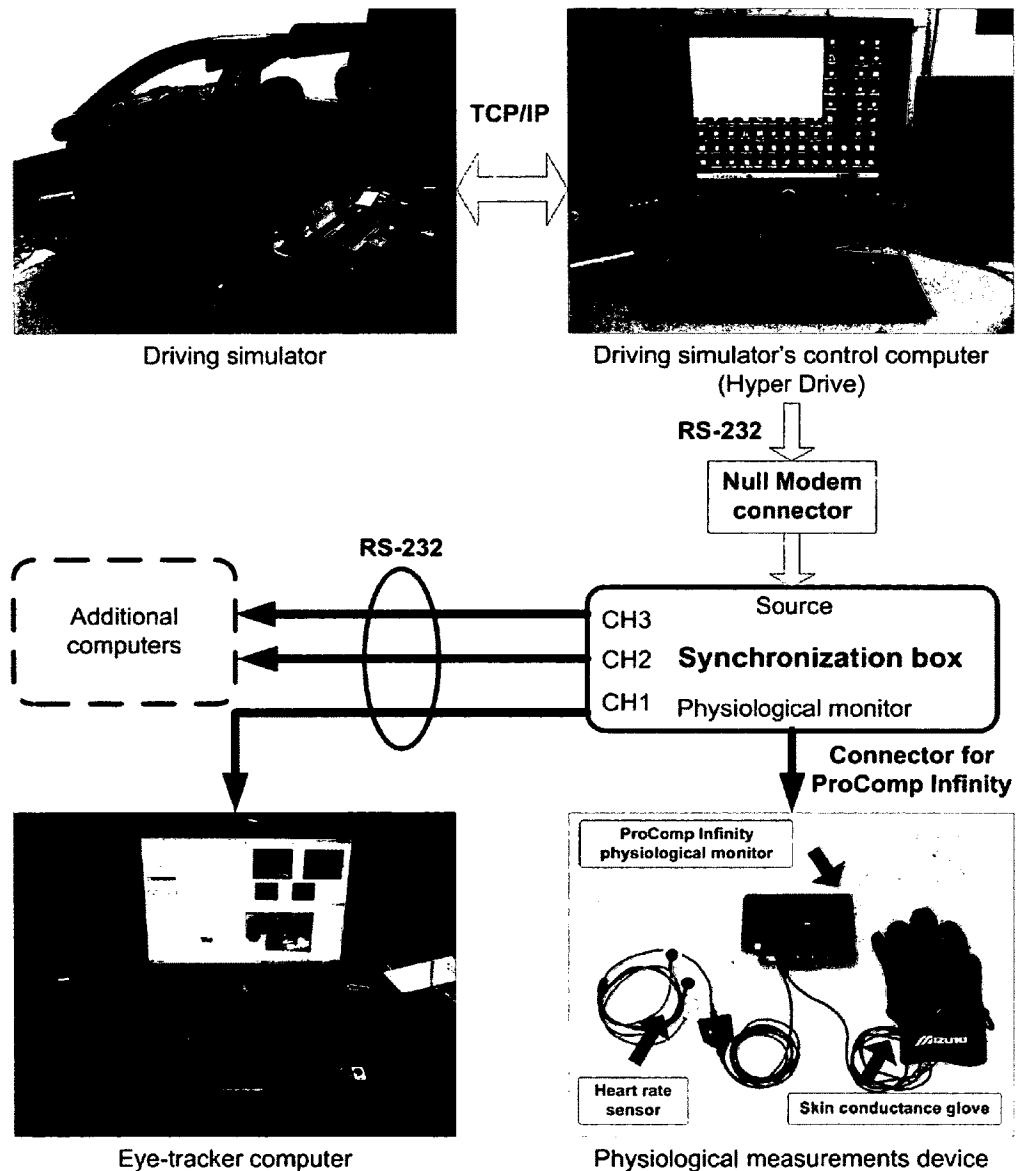


Figure A.3 Synchronizing experimental equipment.

A.1 Hardware Setup

The hardware setup is shown in Figure A.3. Since the control of the experiment is performed from the HyperDrive computer, we decided to use it as a “host” which would send the synchronization messages to other computers and/or equipment involved in the experiment. The main part which connects HyperDrive with other equipment is the synchronization box (Figure A.4).

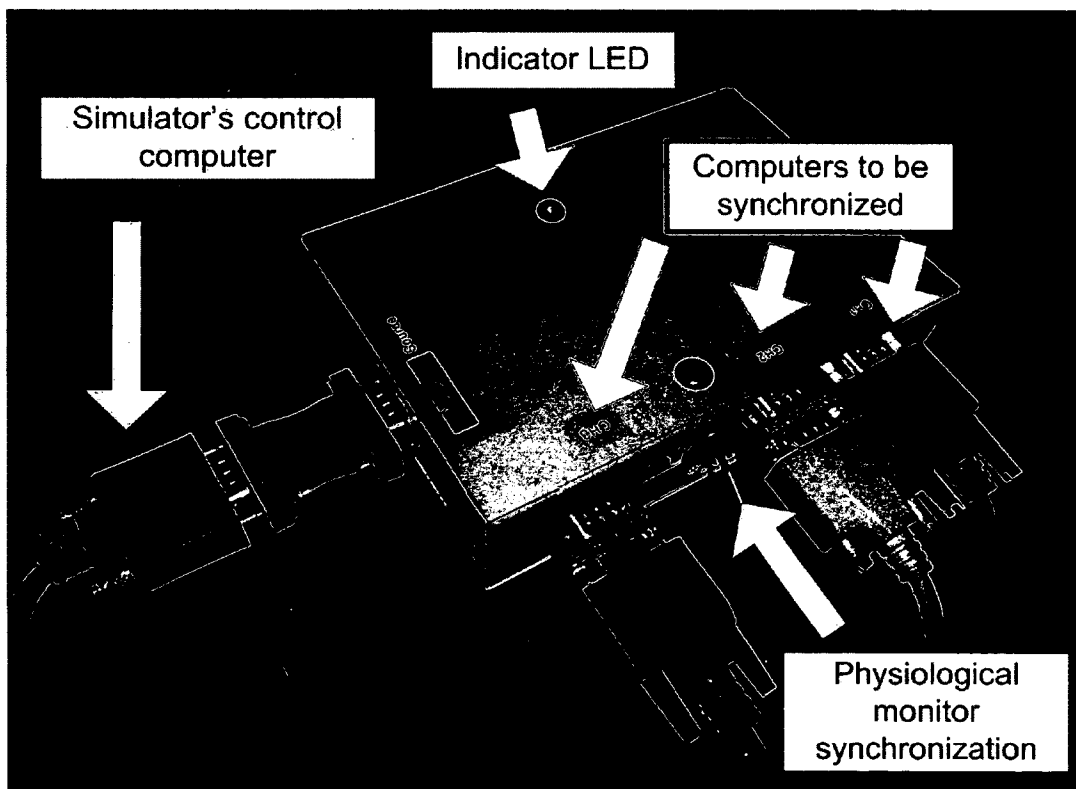


Figure A.4 Synchronization box.

The synchronization box allows HyperDrive to communicate with other computers using the serial RS-232 connector. HyperDrive should be connected (through a null modem converter) to the “Source” terminal, since it is the origin of all the synchronization messages. Up to three computers can be synchronized simultaneously and they should be connected to the terminals labeled “CH1” to “CH3.” Additionally, the

synchronization box allows synchronizing one ProComp Infinity physiological monitor, whose “H” port should be connected to the synchronization box. This connector is specifically designed for ProComp Infinity and cannot be used with other physiological monitors directly. However, the same synchronization principle can be applied with other monitors as well. Finally, an LED indicator is used for a visual confirmation that a synchronization signal has been sent. It stays illuminated as long as the DTR line is set to high on the RS-232 (more details about this functionality will follow in Section A.3). Figure A.5 shows the inside view of the synchronization box, while Figure A.6 shows its detailed schematic.

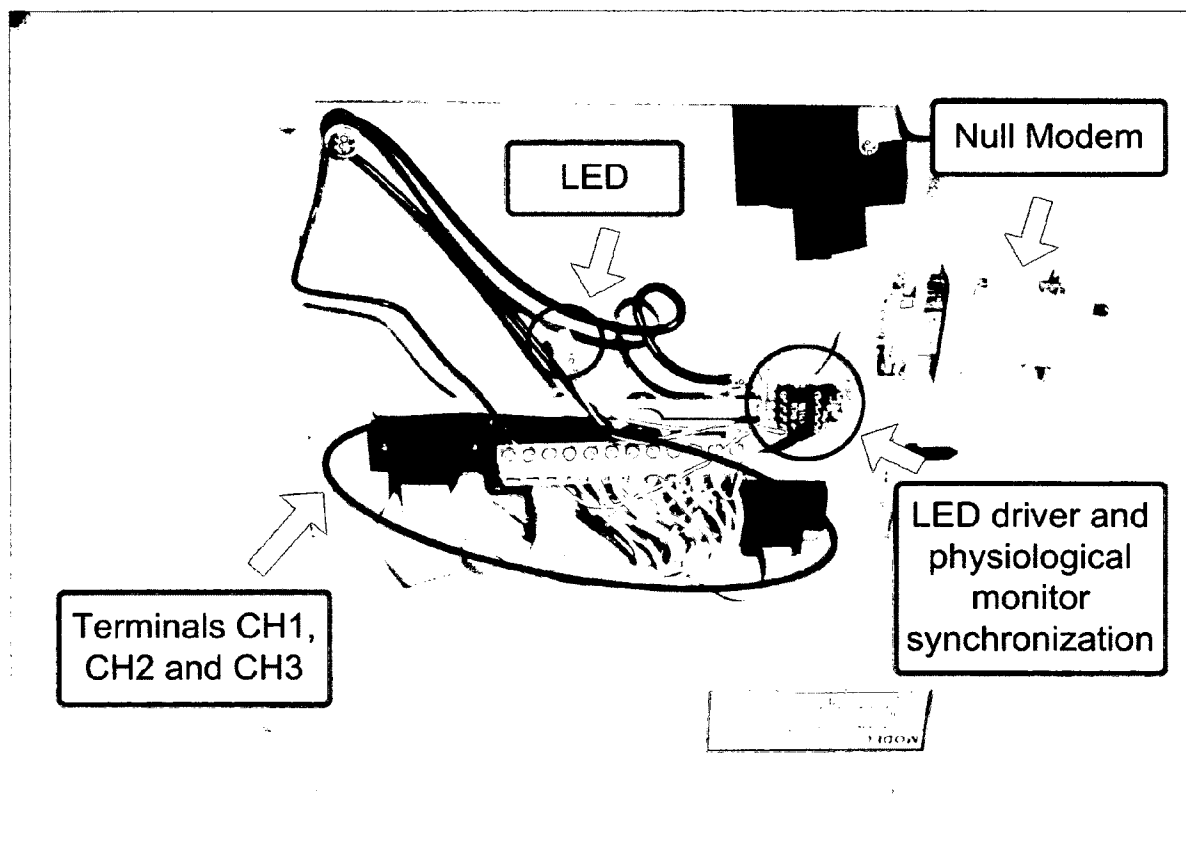


Figure A.5 Inside view of the synchronization box.

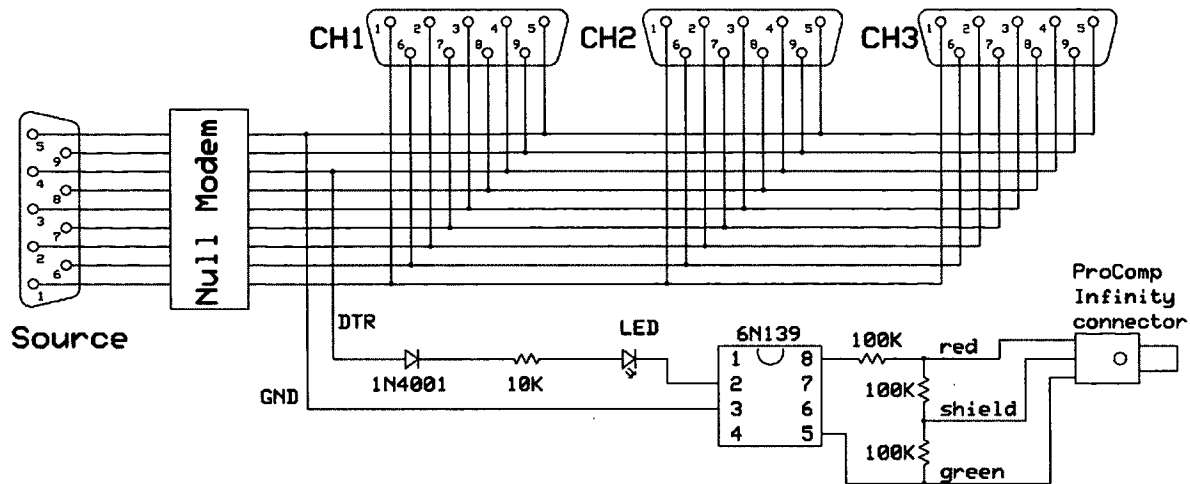


Figure A.6 Schematic of the synchronization box.

As we can see in Figure A.6, the synchronization signals are routed from the “source” computer (in our case HyperDrive computer) through the null modem converter to up to three computers connected to terminals CH1, CH2 and CH3 (in our case eye-tracking computers). Through the indicator LED the “DTR” line is connected to the optoisolator 6N139, which electrically isolates the physiological monitor from the rest of the system (which is required by medical safety standards). The optoisolator plays a role of a switch which is closed when DTR line is high and opened when DTR is low. The rest of the schematic describes the customized connection with ProComp Infinity. The “red,” “shield” and “green” labels indicate the specific wires in the ProComp cable that should be connected to the circuit. If a physiological monitor from another manufacturer is desired to be used only this part of the circuit should be modified.

A.2 Software Setup

The software side of the synchronization is established through an in-house made application called SymConnect (Figure A.7).

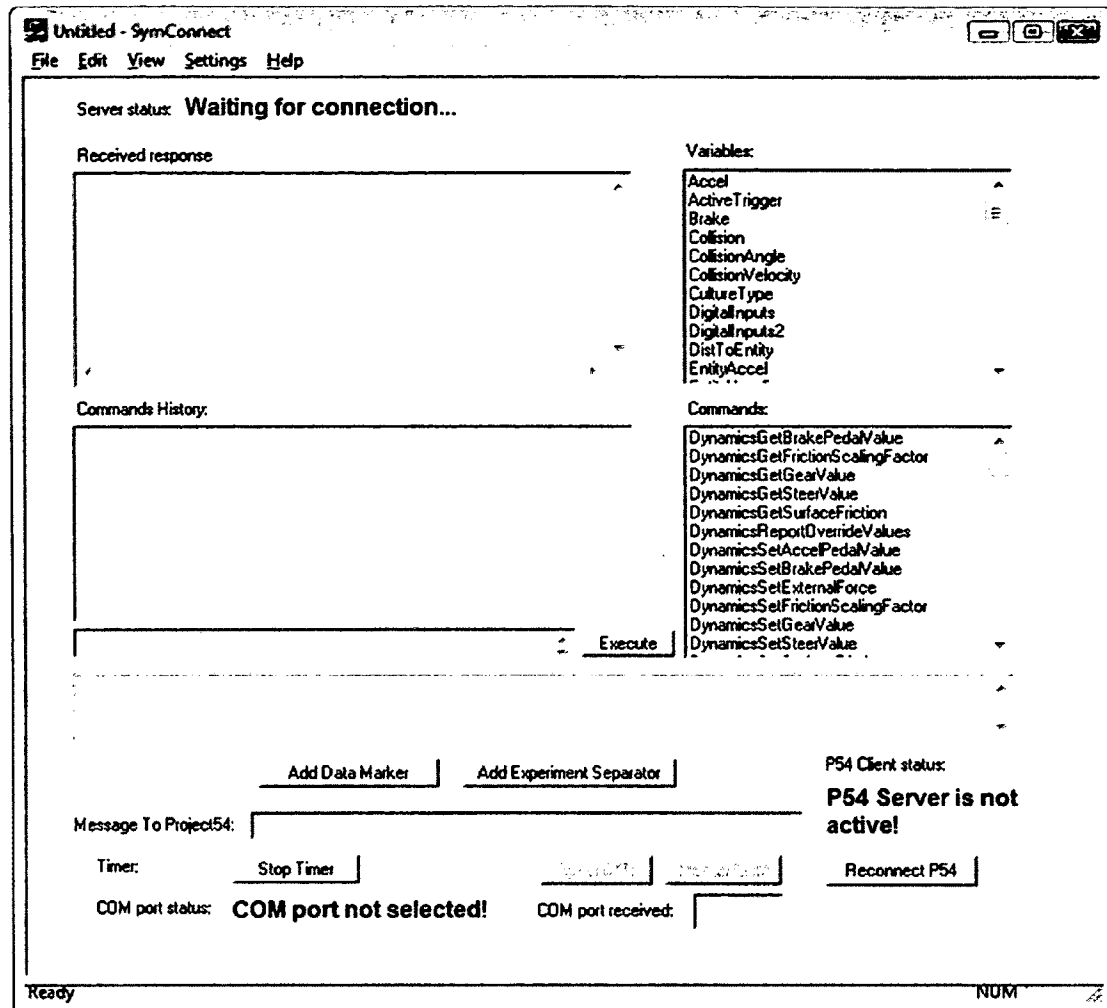


Figure A.7 SymConnect's main window.

SymConnect is a multipurpose application which is used for various tasks, such as synchronization between the driving simulator, eye-tracker(s) and physiological monitor, sending and receiving commands and data between the driving simulator and the Project54 application. Each of these procedures will be explained in the following sections. The source code for SymConnect is under versioning control (Tortoise SVN must be installed and a valid account has to exist in order to access the files) and can be found at this address: <http://pc20m229.unh.edu/svnrepos/hyperdrive/Automation/Controller/SimConnect>. The

following paragraphs will provide explanations of different functionalities which are commonly used in our experiments.

A.2.1 Configuration Files

SymConnect has two configuration files. They are needed to properly set up the connections with the simulator and Project54. The “*configsim.txt*” file configures the TCP/IP communication with the simulator and contains two lines: the first line is the port number that SymConnect listens to, while the second line is the IP address of the local computer (the one running SymConnect) on the simulator’s local network. These numbers can be changed directly from the above file or from within SymConnect by activating “Port and IP” dialog (Figure A.8) in the “Settings” menu (Figure A.9). In both cases, the changes take effect after SymConnect is restarted.

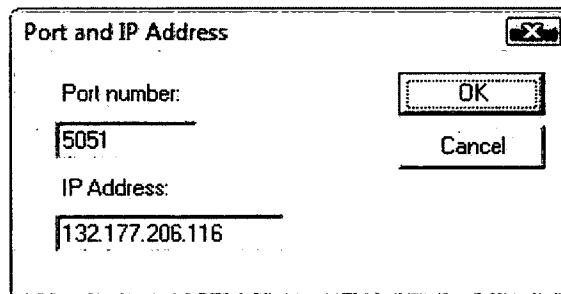


Figure A.8 “Port and IP” dialog.

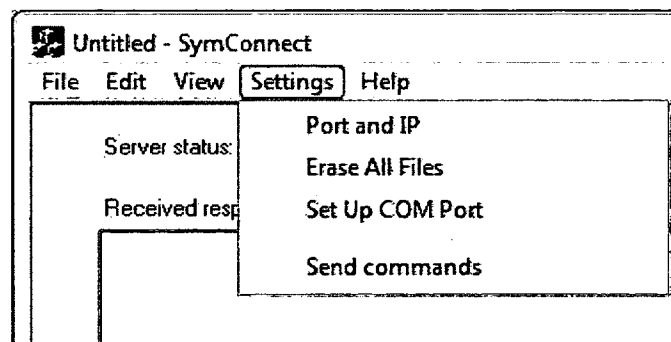


Figure A.9 “Settings” menu.

The “*configP54.txt*” is used to configure the UDP communication with Project54 application and contains the IP address of the computer which is running Project54.

A.2.2 Log Files

Log files record all the activity inside SymConnect. Each time SymConnect is started, all log files are appended with a time stamp which contains time and date on the local computer. This way, the old data is always preserved and can be easily distinguished from the new data. The most important log file is called “*measurements.txt*” and it contains data received from and commands sent to the driving simulator as well as sync signals used for synchronizing all the equipment involved in the experiment. Another log file is “*P54Clicks.txt*” which stores commands sent to the Project54 application (see “Send Commands” option under Section A.2.5).

A.2.3 Establishing Communication between SymConnect and Driving Simulator

In order to establish the communication between SymConnect and the driving simulator a script called “*SymConnect2.tcl*” (can be found here: <http://pc20m229.unh.edu/svnrepos/hyperdrive/trunk/includes/SymConnect2.tcl>) must be included in a desired scenario and invoked from the simulator’s init script. This is done from the Hyper Drive application (running on the HyperDrive computer), which is used for designing scenarios. The code should be invoked at the beginning of the init script using the following syntax:

SymConnect2 show_debug_messages sampling_frequency send_to_SymConnect

If the first parameter is 1, debug messages will be displayed on the screen (default is 0). The second parameter is the frequency with which the data will be sent and received from SymConnect (default is 60 Hz). The last parameter determines if the data should be sent (1) from the simulator to SymConnect or not (0, which is a default value). The data that can be received from the simulator can include values of various variables, confirmations of completed actions, and so on. They are all stored in the “*measurements.txt*” log file. Similarly, various commands can be sent to the simulator from SymConnect. The communication is based on TCP/IP where SymConnect plays the role of the server, while the simulator is the client. Thus, SymConnect must be started *before* a desired scenario (with “*SymConnect2.tcl*” code properly included and invoked, of course) is activated on the simulator. Since the simulator can establish only one connection at any given time, only *one* instance of SymConnect can be running on any local computer (for example, HyperDrive). This is enforced by creating a dummy file called “*SymConnect.lock*” in the root of the local computer. The existence of this file is checked each time SymConnect is started and in case of its existence a warning message will be displayed preventing another instance of SymConnect from starting. Upon SymConnect’s closure, the file is deleted. If SymConnect does not close properly, the lock file needs to be deleted before SymConnect can be started again.

One important variable which is defined in “*SymConnect2.tcl*” script is called “*::sync_pulse*”. This is a global, integer variable which is incremented each time a sync signal is sent to the simulator from SymConnect. By analyzing the contents of this variable from within the simulator’s code, it is possible to determine both when and how

many sync signals have been received. The contents of this variable are stored in the “SyncPulse” column inside the driving simulator’s data collection.

A.2.4 Main Window Options

Figure A.10 depicts SymConnect’s main window.

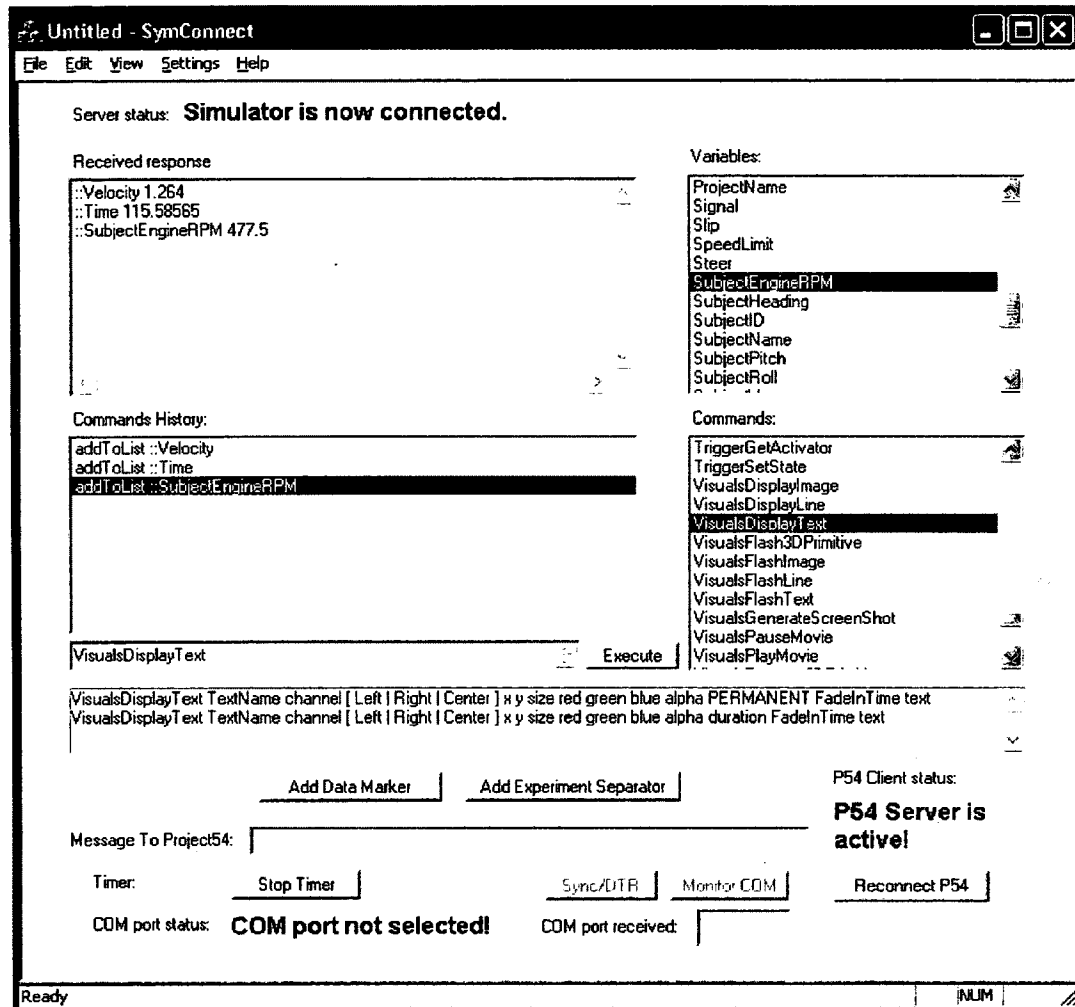


Figure A.10 SymConnect window while sending commands to the simulator.

The main window has multiple regions of interest:

1. “Server status” indicates whether the simulator is connected to SymConnect. If no connection is currently active the text “Waiting for connection...” is displayed. If

the connection is successfully established, the text “Simulator is now connected” is displayed (see Figure A.10). If the connection is not active and a command is sent to the simulator, the text “Socket error while sending packets!” is displayed indicating the inability to send the command.

2. “Received response” displays values of driving variables of interest and any messages received from the simulator (see Figure A.10). All values presented here are logged in “*measurements.txt*” (which can be found inside SymConnect’s local folder) together with the local time when each piece of information is received.
3. “Commands history” displays commands sent to the simulator as well as the sync signals sent to all the equipment used in the experiment (see Figure A.10).
4. “Variables” list displays all the variables that can be obtained from the driving simulator, such as velocity, lane position, and so on. The “Variables” list is automatically populated when SymConnect is started. Its contents are located inside the file called “*variables.txt*.”
5. “Commands” list displays all the available commands that can be sent to the simulator. This includes commands predefined by the simulator software, but also user-defined custom commands can be added. This list is populated during SymConnect’s startup from the file called “*commands.txt*.”
6. Commands text field (located below the “Command history” in Figure A.10) is intended for sending commands to the simulator. The commands can be typed in manually or invoked from the “Variables” and “Commands” lists. By double-

clicking on any of the variables inside the “Variables” list, a special command (called “addToList”) is added to the commands text field. For example, if we double click on variable “Accel” the command “addToList ::Accel” will be added to the commands text field. After pressing the “Execute” button, the command is sent to the simulator and also written inside the “Commands history” field and inside the “*measurements.txt*” file. This command is then received by the “*SymConnect2.tcl*” code on the simulator and added to an internal list of variables whose values are selected to be sent to SymConnect. If enabled within the simulation (by invoking “*SymConnect2.tcl*” code with the *send_to_SymConnect* parameter set to 1), the simulator will start sending the values (at the frequency selected in the *sampling_frequency* parameter) of the selected variable(s) to SymConnect. Figure A.10 gives an example of how the received response looks like in case of three variables: Velocity, Time and SubjectEngineRPM.

Similarly, double-clicking on any command in the “Commands” list inserts it in the commands text field (such as “VisualsDisplayText” command shown in Figure A.10). If available, the selected command’s description is also displayed in the field located below the commands text field. The descriptions are located in “*com_descriptions.txt*” file. New commands and descriptions can simply be added by editing the above files. One specialized command that can be issued through the commands text field is “*Set value.*” This command is handled by “*SymConnect2.tcl*” file, which assigns the parameter “value” to a global variable called “*::test_variable.*” The value of this variable can be checked periodically within the simulation and then acted upon as desired. For example, if the value of

“::test_variable” equals a value of interest, the simulation can display a message on the screen or initiate an event in the simulated environment.

7. “Add Data Marker” and “Add Experiment Separator” are textual indicators used for providing reference points in the “*measurements.txt*” file. They result in adding textual lines which contain phrases “Data Marker” and “Experiment Separator,” respectively. These indicators may be useful if the experimenter wants to indicate when an event of interest occurs (Data Marker) or when the experiment should be separated into individual runs (Experiment Separator). Both indicators are prefixed with time stamps in the log file and besides different titles no other difference exists between those.
8. “P54 Client Status” is a simple indicator which confirms whether Project54 application is running or not. The connection between SymConnect and Project54 is based on the UDP protocol, which is not as strict as TCP/IP. Thus, the connection does not have to be established formally, but rather Project54 will receive any messages sent to it at any time. If Project54 application is started before SymConnect, it will be detected and “P54 Server is active!” message will be displayed (Figure A.10). If this is not the case, “P54 Server is not active!” message will be displayed (Figure A.7). After Project54 is started, pressing “Reconnect P54” button should establish the connection.
9. “Message To Project54” text field allows sending commands to any application inside Project54 [38;145]. The syntax is as follows: “;to_app;;message_string.” The “to_app” parameter indicates the name of the application inside Project54 that should process the message specified in “message_string.” This functionality

is most often used for invoking Project54's text-to-speech engine (handled by the Project54's "speechio" application). For example, if we want the computer to say "hello world," we would issue the following command: ";speechio;;SAYTHIS hello world" (Figure A.11).

Message To Project54: `;speechio;;SAYTHIS hello world`

Figure A.11 Sending a message to Project54's speechio application.

10. "Timer" field counts the number of seconds elapsed since "Add Data Marker," "Add Experiment Separator" or "Sync/DTR" (the discussion of this functionality will be provided in Section A.3) buttons are pressed. Pressing "Stop Timer" button stops the timer and resets it to zero. This functionality may be useful when the experimenter needs to activate desired events manually.
11. The main window also contains the following elements: buttons "Sync/DTR" and "Monitor COM," and indicator text fields "COM port status" and "COM port received." Since these options are used in the process of data synchronization, we will postpone the discussion of their usage until Section A.3.

A.2.5 Settings Menu Options

The "Settings" menu (see Figure A.9) provides multiple important options which are used both for configuring SymConnect and providing additional functionality:

1. "Erase All Files" option clears the contents of *all* log files. This option should be used cautiously, since the erased data cannot be recovered.

2. "Send commands" is a useful option which enables the experimenter to send pre-scripted commands to the simulator or Project54 (Figure A.12 presents the corresponding dialog window).

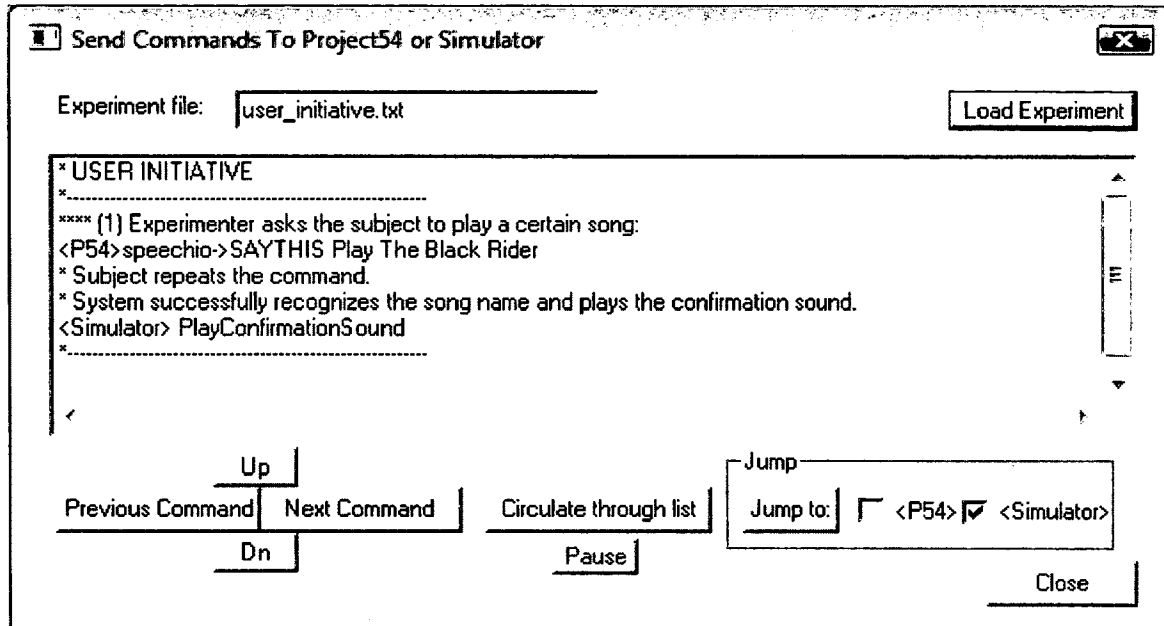


Figure A.12 Send commands window.

The user should first prepare the desired commands in an ordinary text file. As an example, let us say that the file is named "user_initiative.txt." Its example contents are depicted in Figure A.13.

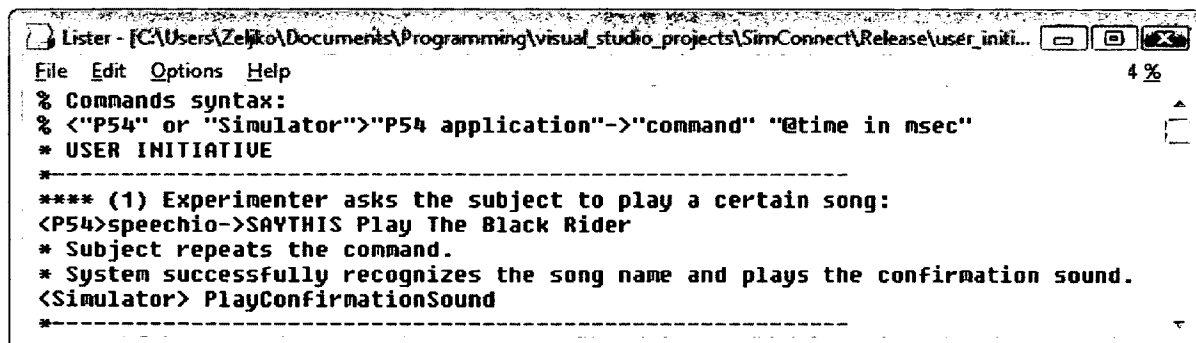


Figure A.13 Contents of the file "user_initiative.txt".

In order to use the commands specified in this file, it must be loaded into “Send commands” window using the “Load Experiment” button or by specifying the name of the file in the “Experiment file” text field. Figure A.12 shows how the window looks like after the file is loaded. In order for the commands to be executed properly, a specific syntax should be followed when creating script files. The syntax is illustrated in Figure A.13 and consists of comments and commands.

Comments: Two types of comments exist and they are distinguished by their prefixing symbol: “%” or “*”. The ones that start with “%” are not displayed in the “Send commands” window and can be used by the experimenter only while compiling the script file (the first two lines in Figure A.13). The second type of comments start with an arbitrary number of asterisks (“*”) and they are displayed in the “Send commands” window. They can be used to place reminders for the experimenter about the necessary steps during the experiment if the commands should be activated manually.

Commands: Based on the desired destination there exist two types of commands: commands for Project54 application and commands for the driving simulator.

Commands that should be sent to Project54 have the following syntax:

<i><P54>application->command</i>

The “application” parameter specifies a desired Project54 application, while “command” represents a desired command, whose syntax depends on the targeted

application. The fifth line in Figure A.13 demonstrates sending a string “Play the Black Rider” to the Project54’s “speechio” application.

Commands intended for the simulator have the following syntax:

<i><Simulator>command</i>

Again, the “command” parameter indicates the name of the command to be sent to the simulator (which, of course, must be handled in the “SymConnect2.tcl” file). If it is desired for SymConnect to traverse the list of the commands automatically (by clicking the “Circulate through list” button, which will be explained in the next paragraph), each command line of the script (either *<Simulator>* or *<P54>*) must be appended with “@time_in_milliseconds” which produces a pause corresponding to the specified interval before the next command can be issued. This way SymConnect knows how long to wait before issuing the next command. Any command can be activated manually by double-clicking on the list. Both *<P54>* and *<Simulator>* commands are logged in “measurements.txt” while a separate log file (“P54Clicks.txt”) is used just for *<P54>* commands.

Now that we have covered the syntax of the script file, we can look into the rest of the interface available in the “Send commands” window. “Up” and “Down” buttons move the selection pointer through the list without executing commands. Conversely, “Previous Command” and “Next Command” buttons move the pointer through the list while executing each command (in doing so, commented lines are skipped). “Jump to:” button moves the selection pointer to the next *<Simulator>* or *<P54>* command (based on the selected check box) and

executes it. “Circulate through list” button enables automatic traversal of the list items and can start from any position of the selection pointer. This option works *only* if each command line specifies a desired time interval in milliseconds before the next command should be executed. If the list traversal is active and the selection pointer lands on a command which does not have the time interval specified, the traversal will stop automatically. Similarly, the traversal will not start unless the selected command specifies a desired time interval. List traversal can be stopped by clicking the “Circulate through list” again; however, the specified time interval has to elapse before issuing new commands. Finally, the list traversal can be paused with the “Pause” button.

3. The last option in the “Settings” menu is “Set Up COM Port.” This option is essential for data synchronization and is used together with multiple controls that reside in the SymConnect’s main window: “COM port status”, “Sync/DTR”, “Monitor COM”, and “COM port received.” Given their importance, the descriptions of the above controls are provided in the next section.

A.3 Data Synchronization

The general logic behind data synchronization is fairly simple: the experimenter issues sync signals through SymConnect, which are then detected by other equipment involved in the experiment and stored in their individual databases. Since the sync signal is received by all the equipment simultaneously, it represents the global reference point from which the beginning of the experiment should be calculated (i.e. a point at which the experiment time should be considered equal to zero).

In order to enable the synchronization, we must first set up SymConnect properly. Since SymConnect uses RS-232 for sending sync signals to other computers and the physiological monitor, we must select an appropriate COM port first. To do that, we choose “Set Up COM Port” option under the “Settings” menu. A dialog depicted in Figure A.14 appears.

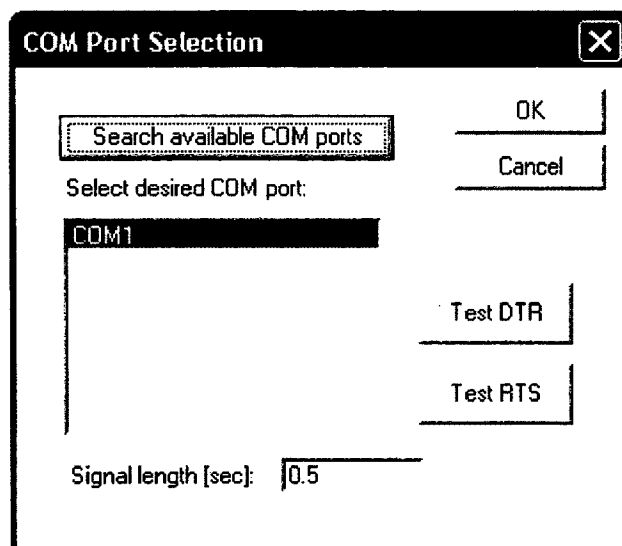


Figure A.14 COM port selection.

Clicking “Search available COM ports” will find all the available COM ports on the local computer and display those in the “Select desired COM port” list. The desired COM port should then be selected from the list. If desired, the experimenter can confirm if the selected port is correct by specifying the duration of the test signal in the “Signal length [sec]” field and clicking the “Test DTR” button. This will produce a DTR signal of the specified duration on the RS-232 connector, which can be visually inspected by observing the indicator LED on the synchronization box (see Figure A.4). By clicking the “OK” button, the selected COM port is saved and opened with the following characteristics of the serial communication: 8 data bits, no parity, 1 stop bit, 1200 baud

and software flow control set to on. After the COM port is successfully opened, two changes occur in the SymConnect's main window (see Figure A.15). First, the "COM port status" indicator changes from "COM port not selected!" to "COM1 is active" which is an indication that the port (COM1 in our case) was successfully initialized and synchronization is possible. And second, the buttons "Sync/DTR" and "Monitor COM" become available. More details about the usage of these buttons will follow shortly.

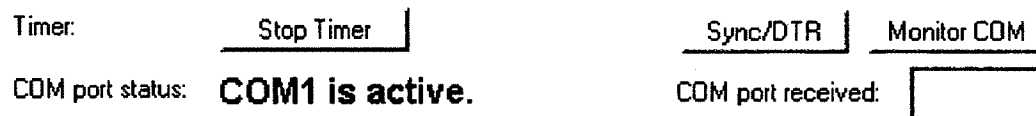


Figure A.15 Synchronization is enabled by activating the COM port.

As we mentioned before, only one instance of SymConnect can be running on any computer. However, multiple instances can be running on separate computers. As a matter of fact, this is even required when the synchronization is preformed with the eye-tracker computer(s) or any other computer. The instance of SymConnect that the experimenter is using for issuing sync signals (from now on, we will refer to it as "main" SymConnect) should be running on the HyperDrive computer (alternatively, it can run on any other computer connected to the simulator's network, with IP addresses correctly specified inside SymConnect and "SymConnect2.tcl"; however, since in our case HyperDrive controls the simulation and retrieves the data, it is natural for it to run the main instance of SymConnect as well).

Sending sync signals is accomplished by clicking the "Sync/DTR" button in the main instance of SymConnect, which invokes a function named "OnBnClickedButtonmanualdtr." This function generates three consecutive sync

signals, one for each device that can be synchronized: eye-tracker (symbol “s”), driving simulator (word “SYNC”) and physiological monitor (high level on DTR line on the RS-232 connector). We tested how fast those sync signals are issued in a typical experimental setting by periodically clicking the “Sync/DTR” button (every 2 seconds) for 2000 times. Each time this button is pressed, SymConnect logs the local time when each of the three sync signals was issued. By comparing those times we determined that the delay between sending those signals was always equal to 0 msec, which indicates that they were indeed sent to their recipients at the same time.

Synchronizing eye-tracker: Since the eye-tracker software cannot directly accept sync signals, it is necessary to run a separate (“secondary”) instance of SymConnect on the eye-tracker’s computer. The setup of this instance is exactly the same as with the main SymConnect, however, the “Monitor COM” button should be activated. Clicking the “Monitor COM” button invokes a function named “OnBnClickedButtonmanualrts.” This function performs multiple actions: disables the “Sync/DTR” button, displays the word “Waiting” (Figure A.16, left) in the “COM port received” text field and starts a thread named “COMCheck,” which puts the application in the “listening” mode, where it waits for the sync signals coming from the selected COM port. What this means is that “COMCheck” waits (in a blocking “Read” call) until some content appears in the buffer of the selected COM port. Specifically, the code checks if the received symbol is “s” which is an indication that the sync signal was received from the main SymConnect. The receipt of the sync signal is indicated by the word “SYNC” in the “COM port received” field (Figure A.16, right). Multiple sync signals can be received from the main instance of SymConnect and all of them are

recorded in the “*measurements.txt*” log file together with their corresponding time stamps. Naturally, sync signals are also recorded inside the main SymConnect’s log file as well. Since the eye-tracker software assigns the local time to each data sample collected in its database, and because it is known when the sync signal is received in the local time (by observing the “*measurements.txt*” log file coming from the local instance of SymConnect), it is possible to pinpoint the exact location in the eye-tracker’s database when the sync signal is received by the eye-tracker computer. This time then becomes the “zero” reference point for the eye-tracker data.



Figure A.16 Waiting for sync signal (left) and signal received (right).

In order to test the time delay that elapses between sending the sync signal from the main SymConnect and receiving it by the secondary SymConnect, we performed a round-trip delay test. In this test a sync signal is sent to the secondary SymConnect and immediately reflected back to the main SymConnect. If we measure the total round-trip time and divide it by two, we can obtain the time delay that is necessary for the secondary SymConnect to receive the sync signal from the main SymConnect (one-way delay). By periodically sending the sync signal (approximately every 2 seconds) for 3000 times, we obtained the following results for the one-way delay: maximum delay 46.5 msec, minimum delay 39 msec, average delay 40 msec and standard deviation of the delay 2.6 msec. Based on this test we can make two conclusions. First, the delay varies very little, which indicates its consistency and reliability. And second, it is much smaller than our data sampling period of 100 msec,

which indicates that it provides much higher precision than necessary for our data collection.

Synchronizing driving simulator: In case of the driving simulator, “SymConnect2.tcl” code handles the synchronization. It listens on the selected port (recall that in this case it is a TCP/IP port) for the received packets from the main SymConnect by periodically (this frequency can be customized, but default is 60 Hz) pooling the contents of the port. If the received packet contains the word “SYNC,” a global variable “::sync_pulse” (defined in “SymConnect2.tcl”) is incremented by 1. Its value is then written in the data collection under the column named “SyncPulse.” This way the number of sync signals is counted, so the value assigned to the “SyncPulse” column reflects both the number and the time when each sync signal is received.

In order to test the time delay that elapses between the sending of the sync signal from the main SymConnect and receiving it by the simulator, we performed the same round-trip delay test as in the case of COM port communication, by periodically (approximately every 2 seconds) sending a total of 2000 sync signals. We obtained the following results for the one-way delay: maximum delay 7.5 msec, minimum delay 0 msec, average delay 6.3 msec, standard deviation of the delay 2.7 msec. Based on these results we can obtain the same conclusions as with the COM communication: the delay is very consistent and reliable and provides significantly higher precision compared to our data collection period of 100 msec.

Synchronizing physiological monitor: Finally, in case of the physiological monitor, main SymConnect raises the DTR line on the RS-232 connector to a high level for the number of seconds specified in the “Signal length [sec]” field in the “COM Port

Selection” dialog window (default value is 0.5 seconds). This change in the voltage level is then transmitted by the synchronization box to the physiological monitor (H slot should be connected to the synchronization box). The change in the voltage level is sampled by the physiological monitor’s A/D converter and recorded in the device’s database, thus allowing precise determination of the zero reference. Our physiological monitor (ProComp Infinity) samples its H port at 256 Hz. Unfortunately, it has no capability to reflect the received sync signal to the origin in order to analyze the one-way delay. However, since all the components involved in synchronizing the physiological monitor are hardware based, we can be fairly certain in assuming that the maximum delay is not larger than $1/256 = 3.9$ msec, which provides much higher precision compared to the sampling period of 100 msec used in collecting the rest of the data.

A.4 Simple Experiment Example

Since all the important data pertaining to SymConnect is collected in “*measurements.txt*” we will look at its contents using a simple experiment. Imagine that we want to synchronize a single eye-tracker computer with the simulator and also to observe three simulator variables in real time through the main instance of SymConnect: Velocity, Time and SubjectEngineRPM. We will assume that the main SymConnect resides on the HyperDrive computer and the secondary SymConnect resides on the eye-tracker computer. Also, the eye-tracker and the HyperDrive computers must be connected to the synchronization box as depicted in Figure A.3. Finally, “*SymConnect2.tcl*” must be properly included and invoked in the desired simulated scenario in order to establish the communication with the main SymConnect on the HyperDrive computer. The following

steps should be completed (steps 1 through 5 are general synchronization steps and should be performed at the beginning of every experiment):

1. Start main SymConnect on the HyperDrive computer and select a desired COM port,
2. Start secondary SymConnect on the eye-tracker computer, select desired COM port and activate “Monitor COM” option,
3. Start eye-tracking,
4. Start a desired simulated scenario and wait until the simulator successfully connects to the main SymConnect (indicated by the string “Simulator is now connected.”),
5. Press the “Sync/DTR” button, which sends the sync signals to all connected equipment. Visual confirmation can be accomplished by observing both the LED indicator on the synchronization box as well as by observing the “COM port received” field in the secondary SymConnect on the eye-tracker computer. Furthermore, the sync signal can be detected inside the simulator’s init script by examining the value of a global variable named “::sync_pulse” which is defined inside the “*SymConnect2.tcl*” script. This variable is used for counting the number of issued sync signals and for initializing the contents of the “SyncPulse” column inside the simulator’s data collection. By detecting changes in the “::sync_pulse” variable we can program various actions to be performed inside the simulator. For example, confirmation messages can be

displayed on the screen, which would give a visual indication when the sync signal is received.

6. (Optional) As per our example, we should double-click on each of the desired variables in the “Variables” list in the main SymConnect followed by the “Execute” button (see Figure A.10).
7. (Optional) If desired, by clicking on the “Data Marker” and “Experiment Separator” buttons it is possible to indicate important parts of the experiment by adding their corresponding markers in the “*measurements.txt*” file.

Figure A.12 shows the abbreviated version of the main SymConnect’s “*measurements.txt*” file after performing the simple experiment described above. Row numbers that can be seen at the beginning of each line are not part of the “*measurements.txt*” file and they were added in order to facilitate the explanation of the file’s contents.

```

1. --- 15:00 Jan.20.12. ---
2. (15:17:00:800 >>) Simulator is Connected and Ready to Receive Commands
3. (15:17:04:805 <<) Remote SYNC
4. (15:17:04:805 <<) Sim SYNC
5. (15:17:04:805 <<) <PM>DTR

6. (15:17:06:339 <<) addToList ::Velocity
7. (15:17:08:839 >>) Velocity 0.000
8. (15:17:09:128 >>) Velocity 0.000

9. (15:17:16:995 <<) addToList ::Time
10. (15:17:17:495 >>) Velocity 0.000
11. (15:17:17:495 >>) Time 24.08381

12. (15:17:24:854 <<) addToList ::SubjectEngineRPM
13. (15:17:25:354 >>) Velocity 0.000
14. (15:17:25:354 >>) Time 31.98397
15. (15:17:25:354 >>) SubjectEngineRPM 477.5

16. (15:17:56:645 >>) Time 63.58461
17. (15:17:56:645 >>) SubjectEngineRPM 477.5
18. (15:17:56:645 >>) Velocity 1.261
19. (15:17:56:745 <<) ExperimentSeparator

20. (15:23:57:182 >>) Time 69.08462
21. (15:23:57:182 >>) SubjectEngineRPM 477.5
22. (15:23:57:182 >>) Velocity 1.261
23. (15:23:58:007 <<) DataMarker

```

Figure A.17 Sample main SymConnect's "measurements.txt" file.

Line 1 indicates the date and time when SymConnect was started. As we can see, each line is preceded with the local time stamp. Received data is symbolically indicated by ">>", while the data sent by SymConnect is indicated by "<<". Line 2 indicates that the simulator successfully established the connection with SymConnect. Lines 3, 4 and 5 indicate when the experimenter pressed the "Sync/DTR" button. By

doing so, the following three sync signals were generated: for the secondary SymConnect on the eye-tracker's computer ("Remote SYNC"), for the driving simulator ("Sim SYNC") and for the physiological measurements monitor ("<PM>DTR"). Line 6 indicates when the experimenter added "Velocity" to be received from the simulator. Similarly, lines 9 and 12 indicate adding "Time" and "SubjectEngineRPM" variables. Immediately after adding each of those variables, the simulator starts periodically (at the frequency specified when invoking "*SymConnect2.tcl*") sending their values to SymConnect. This can be seen in the lines containing the names of the above variables. Finally, lines 19 and 23 indicate when the experimenter pressed the "Experiment Separator" and "Data Marker" buttons, respectively.

Finally, Figure A.18 shows how the secondary SymConnect's "*measurements.txt*" file looks like after concluding the experiment. As we can see, it contains one sync signal and indicates the local time when it was sent from the main SymConnect.

```
--- 15:02 Jan.20.12. ---  
(15:17:40:965 >>) Remote SYNC
```

Figure A.18 Sample secondary SymConnect's "measurements.txt" file.

A.5 Synchronizing Audio Recordings

The previous sections demonstrated how to synchronize data collections located on multiple computers which are commonly involved in driving simulator experiments. Besides log files, it may be of interest to synchronize audio recordings as well. Namely, in case of driving simulator studies which explore auditory interactions it

is useful to record participant's utterances. These audio recordings can be used in post-processing to analyze various aspects of conversation, such as pauses, number of words uttered per minute, word choices, interruptions, etc. For this to be possible, it is necessary to synchronize the audio recordings with the rest of the equipment.

One possible implementation is to introduce an audible signal ("beep") into the audio recording (this has to be done through mixing, which we do not describe here) at the same time the sync signal is issued from the main SymConnect (as a reminder, this sync signal is received by all the equipment involved in the experiment, thus representing a "global reference" which indicates the beginning of the experiment). As explained before, the occurrence of the sync signal can be detected in the simulator's script by periodically examining if the value of the global variable "::sync_pulse" (defined in "*SymConnect2.tcl*") has changed. This variable reflects the number of sync signals received from SymConnect, so by comparing the new value with the saved old value we can decide when the signal actually appeared.

The simulator provides one digital output signal which can be controlled from within the init script. This signal is used to control the dashboard light and its status can be set to *on* or *off* using the predefined command "*VehicleSetDashLight*". When the sync signal is detected, the following command should be executed inside the init script for a predefined amount of time (0.5 seconds appears to be enough for easy detection in an audio file):

<i>VehicleSetDashLight On</i>

After the predefined time period elapses, the signal should be turned off using the same command, but with the parameter set to "*Off*". The result of these actions is a digital

signal with a value of 0 before the sync signal is received, V_{cc} for 0.5 seconds after the sync signal, followed by 0 again (see Figure A.19).

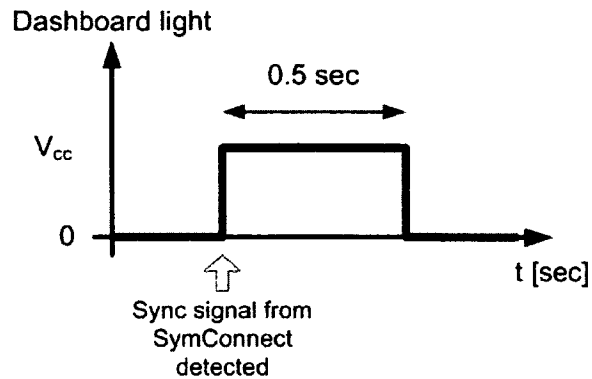


Figure A.19 Dashboard light signal.

Now that we have a physical signal (dashboard light) from the simulator, we can use it for synchronizing an audio signal. For this purpose, a simple astable multivibrator circuit was designed (Figure A.20).

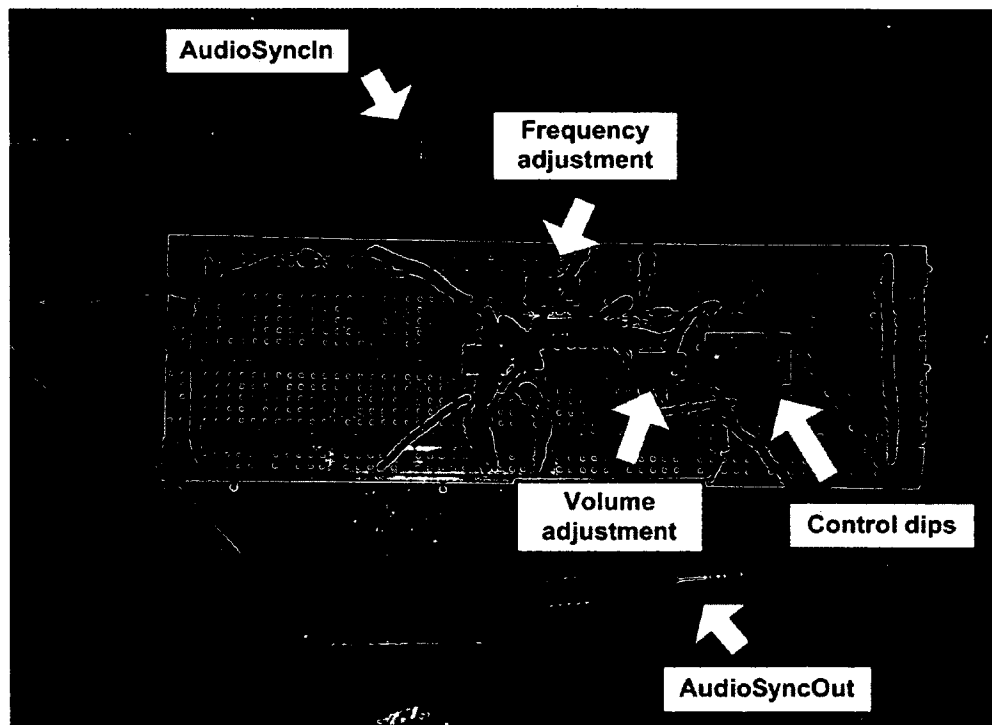


Figure A.20 Breadboard with the astable multivibrator circuit.

[illegible]

Since the dashboard light signal turns on by default whenever the simulator's cabin is first powered on, it is necessary to turn it off before connecting the

“AudioSyncOut” connector to an audio recording device. Otherwise, the circuit will continuously produce the synchronization sound. This can be accomplished by starting any simulation which has “*VehicleSetDashLight Off*” command specified at the beginning of the init script. This will set the dashboard light signal to 0 and it will remain 0 as long as the simulator cabin remains turned on or the experimenter manually toggles the signal to 1. Once this is done, dip switches 1 and 2 can be moved to their “up” positions, which will enable the output.

APPENDIX B

EXPERIMENTAL APPARATUS

This appendix provides descriptions of the equipment employed in various studies throughout this dissertation, specifically driving simulator, eye-tracker and physiological monitor.

B.1 Driving Simulator

The experiments described in this dissertation were performed in a Drive Safety DS-600c Research Simulator [110]. It is a high-fidelity driving simulator (Figure B.22), with the following characteristics:

1. 5 visual channels: 3 front channels which make a 180° field of view screen, 2 side mirrors and 1 rear-view mirror,
2. full-width car cabin (Ford Focus) with realistic vehicle dynamics (vibrations and sounds): motion platform providing inertial cues through a combination of $\pm 2.5^\circ$ pitch and 5 inch longitudinal movement, haptic feedback on the steering wheel, gas and brake pedals and a fully functional dashboard with the corresponding instrumentation.



Figure B.22 High fidelity driving simulator.

The simulated environments (scenarios) are designed using a graphical user interface (GUI), which supports various surroundings, such as urban, rural, residential, suburban, industrial and commercial. The system possesses an extensive library of different road types, intersections, road signs, vehicles, and so on. There is a support for fully automated ambient vehicles which obey the traffic laws, signs, traffic lights and adjust their decisions based on the human behavior. Using the Tcl/Tk scripting language, the researchers can predefine the behavior of the objects of interest (such as pedestrians or vehicles), thus making it possible to simulate a wide variety of traffic situations. Furthermore, Tcl/Tk allows communication over the local area network (LAN), which makes it possible to exchange data between the simulator and a third party software in real time while the simulation is running. As we had a chance to see in Appendix A this capability has been used in our driving simulator studies for the purpose of

synchronization with the external data collections or issuing commands to the simulation system. The simulator provides a wide range of standard driving performance data (such as lane position, velocity, acceleration, steering wheel angle, etc.) at selectable sampling rates of up to 60 Hz.

B.2 Eye-tracker

Most of our studies employed an eye-tracker for analyzing drivers' visual attention. We used a Seeing Machines [146] faceLab 5.02 stereoscopic remote eye-tracker. The eye-tracker was mounted in front of the driver on top of the dashboard (see Figure B.23). As we can see in Figure B.23, the eye-tracker consists of two cameras and an infrared illuminator, which produces a reflection in subject's eyes that the software is using for tracking the eye movement.

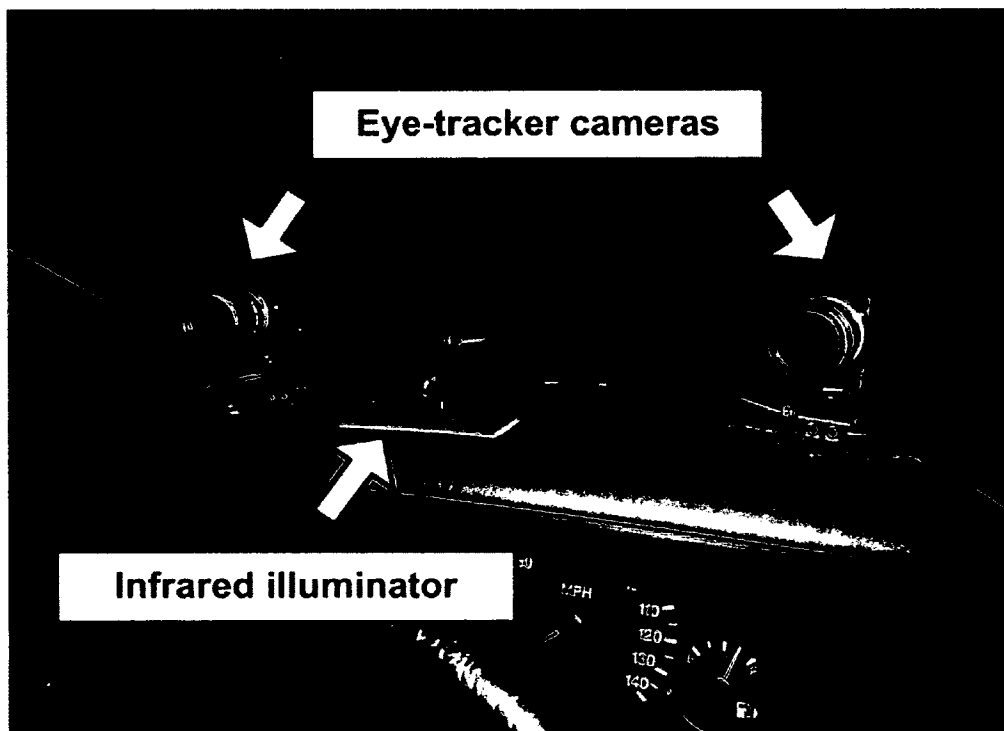


Figure B.23 Eye-tracker mounted on top of the dashboard.

Figure B.24 shows a view of the participant as seen by the eye-tracker. The green vectors coming out from the participant's eyes indicate the direction of the gaze, while the red vector that can be seen between the eyes indicates the direction of the head. The caption "FrontScreen" indicates that the gaze is directed towards the simulator's front screen. The number "00000354" shows the number of the current frame for which the calculations are performed. This information is very useful when manual correction of the eye-tracker data is necessary (for example, when the eye-tracker loses tracking due to the subject obstructing the view of the cameras with hands or when turning the head too far to either side).

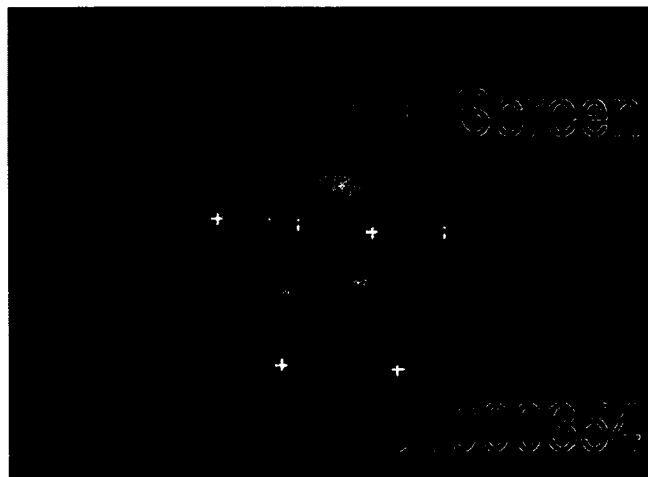


Figure B.24 A view of the participant as seen by the eye-tracker.

The eye-tracker software provides various data corresponding to the eye and head movements at the rate of up to 60 Hz. Some of the data we were interested the most in our studies included objects that a participant is focusing on (used in post-experiment analyses to calculate the PDT on the road ahead, glance duration and glance frequency away from the road, number of glances, etc.) and pupil diameter. As described in Chapter 2, pupil diameter may be useful in describing the overall experienced cognitive load. The

eye-tracking software provides an estimate of pupil diameter based on an ellipse fitting algorithm. Figure B.25 shows how the fitted pupil diameter looks like (green ellipses).

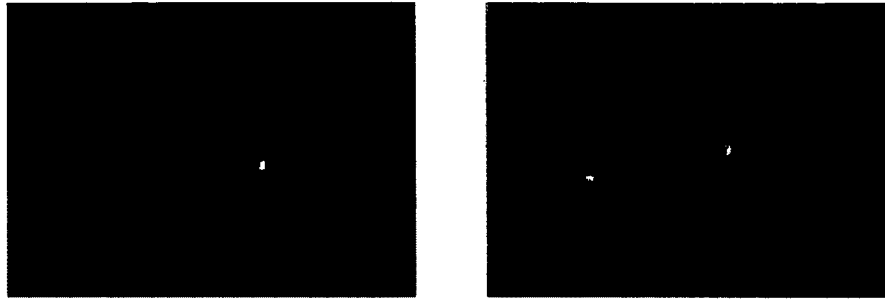


Figure B.25 Fitted pupil diameter.

Finally, the software allows defining a virtual car cabin with all the objects of interest specified with respect to their size and spatial location. Figure B.26 shows how the virtual model looks like in the case of our driving simulator.

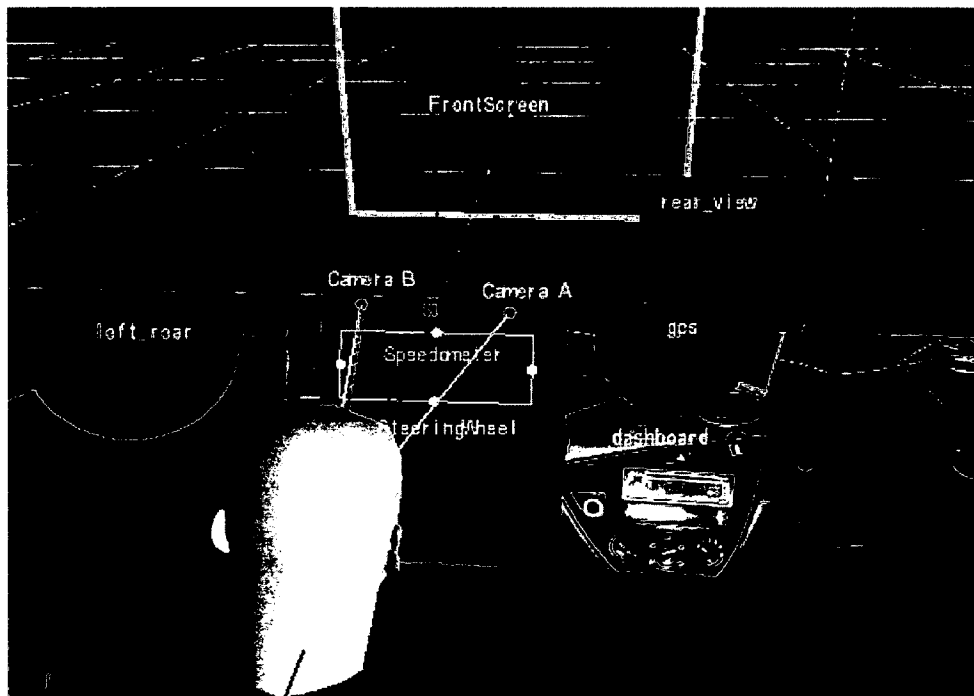


Figure B.26 Virtual model of the car cabin.

There are multiple objects in the model, for instance, front screen, rear view mirror, GPS, speedometer, etc. The yellow avatar simulates the participant's head. There are two vectors which protrude from the avatar: green and red. The green vector indicates the direction of the subject's gaze, while the intersection between this vector and any object in the model indicates the object that the participant is looking at (green dot on the "FrontScreen" in Figure B.26). Similarly, the red vector shows the direction of the head.

B.3 Physiological Monitor

Figure B.27 shows the physiological monitor which was employed in our studies. It is a Thought Technology ProComp Infinity [147] physiological monitor.

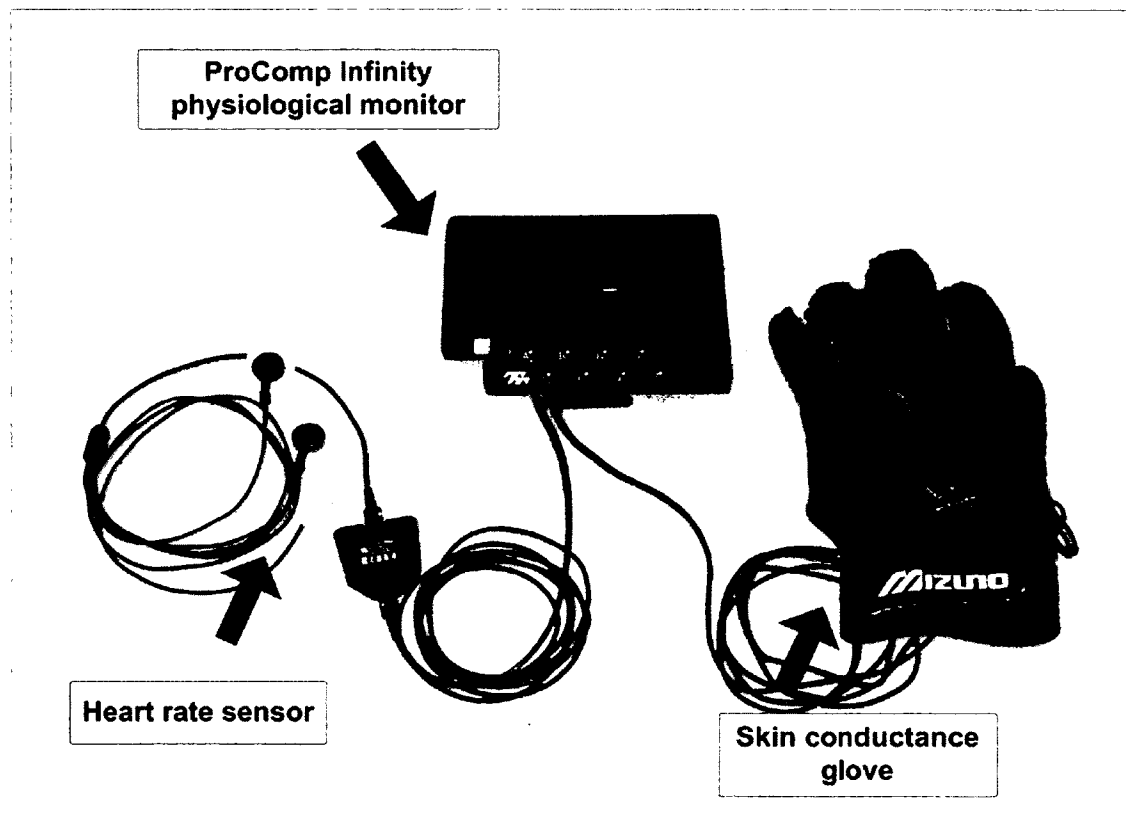


Figure B.27 Physiological monitor and the corresponding sensors for skin conductance and heart rate.

ProComp Infinity has 8 channels of which two are sampled at 2048 Hz and 6 are sampled at 256 Hz. As we can see in Figure B.27 we used two channels: one for the heart rate sensor (channel A, 2048 Hz) and one for the skin conductance sensor (channel E, 256 Hz). If we recall from Appendix A, we also used one additional channel for data synchronization (channel H, 256 Hz). The sampled data can be recorded directly on a computer through an optical cable or can be stored locally on an SD card.

The original skin conductance sensor consisted of two electrode straps (Figure B.28, right) which should be mounted on the tips of the fingers. However, there are two reasons which made this solution unsatisfactory in the driving simulator. First, since the participants were required to operate the steering wheel, the cables would often get entangled. This increased the obtrusiveness of the sensor and made the driving experience unnatural. And second, operating the steering wheel often resulted in the wires detaching from the electrode straps (as we can see on the right of Figure B.28, the wires are attached with the snap-on buttons). Given these problems I decided to embed the electrodes in a glove (Figure B.28, left), which solved both of the above problems.

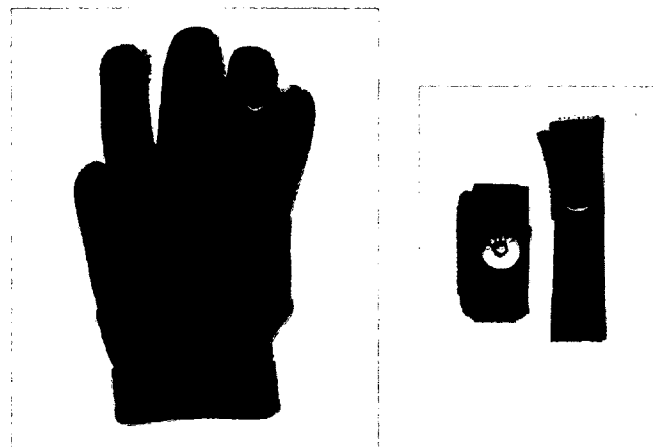


Figure B.28 Skin conductance electrodes embedded in a glove (left) and electrode straps (right).

B.4 Institutional Review Board Form

UNIVERSITY OF NEW HAMPSHIRE

INSTITUTIONAL REVIEW BOARD FOR THE PROTECTION OF HUMAN SUBJECTS IN RESEARCH

Purpose:

This research is funded by the National Institute of Justice (NIJ). The purpose of this research is to assist in the development of speech user interfaces as well as other user interfaces for mobile environments such as vehicles and handheld computers. Another goal is to develop specific applications for mobile environments, specifically for vehicles and for places where people use handheld computers.

Procedure:

You will be asked to interact with the Project54 system running on a PC and/or on a handheld computer. You may also be asked to perform a physical task, such as operating a driving simulator. The Project54 system will record your speech, and/or your interactions with the GUI and/or your interactions with original hardware interfaces, and/or data generated by electronic devices that you interact with and/or data generated by electronic devices that the Project54 system interacts with. The recording will require no special steps on your part. You will also be asked to respond to questionnaires that will ask for personal information and feedback about the experiment.

You will be asked to interact with a PC and/or on a handheld computer and/or other electronic devices. You may also be asked to perform a physical task, such as operating a driving simulator. We will create audio and/or video recordings of your

interactions. We will also record your interactions with the computer's GUI and/or your interactions with other hardware interfaces, and/or data generated by the computer and/or by the electronic devices. We may also record physiological measurements from sensors attached to your body (e.g. temperature, electrocardiogram, skin conductance sensors), and/or sensors in your environment (e.g. pressure sensors on objects in your environment, gaze and head position trackers). You will also be asked to respond to questionnaires that will ask for personal information and feedback about the experiment.

Data generated in this research will be saved for use in future research. A unique ID will be assigned to you. The unique ID will be of the form "User #xx", where xx is the number assigned to you. It will be used to label your data, along with your age, gender, characteristics of your speech, your experience in working with computers or the Project54 system and any questionnaires you fill out. The data will be stored for future use in our research; there is no set date for destruction of the data, and it may be kept for an unlimited duration. Your identity will not be tied to the data in any way other than to the video data, if such data is created, since video data may visually identify you. Video data may be generated by stand-alone video cameras and by cameras that are part of a gaze and head tracker. In this document we are asking for your consent to participate in our study and to share the non-video data with researchers from other institutions. Separately we ask for your consent to share video data with researchers from other institutions, to include still shots from videos in scientific publications and technical reports, as well as to show video data at conferences and similar meetings. Finally, we also ask for your consent to share video data with the public by posting video clips, or

still shots from the clips, online (on sites such as Flickr or YouTube), or by including them in printed publications.

The only risks associated with this research are the potential of skin irritation from sensors attached to your body and the potential for motion sickness if operating a driving simulator. There should be no aftereffects of this research upon you. You will be compensated at approximately \$___/hour for your effort. Your compensation may be in the form of a check or in the form of a gift certificate or in the form of a software license (provided by Microsoft). You may have to fill out a W-9 form. Checks will be mailed by UNH. Your compensation may be reported to the IRS.

1. You understand that the use of human subjects in this project has been approved by the UNH Institutional Review Board for the Protection of Human Subjects in Research.

2. You understand the scope, aims, and purposes of this research project and the procedures to be followed and the expected duration of your participation.

3. You have received a description of any reasonable foreseeable risks or discomforts associated with being a subject in this research, have had them explained to you, and understand them.

4. You have received a description of any potential benefits that may be accrued from this research and understand how they may affect you or others.

5. The investigator seeks to maintain the confidentiality of all data and records associated with your participation in this research. You should understand, however, there may be rare instances when, in order to comply with policy, regulations or

laws, the investigator is required to share personally-identifiable information for research-related purposes (e.g., officials at the University of New Hampshire, designees of the sponsor(s), and/or regulatory and oversight government agencies may require access to research data in order to investigate a complaint about the conduct of the research). Personally-identifiable information will not be released for non-research purposes without your prior consent.

6. You understand that your consent to participate in this research is entirely voluntary, and that your refusal to participate will involve no prejudice, penalty or loss of benefits to which you would otherwise be entitled.

7. You further understand that if you consent to participate, you may discontinue your participation at any time without prejudice, penalty, or loss of benefits to which you would otherwise be entitled.

8. You confirm that no coercion of any kind was used in seeking your participation in this research project.

9. You understand that if you have any questions pertaining to the research you can call Dr. Andrew Kun at 603-862-4175 and be given the opportunity to discuss them. If you have questions pertaining to your rights as a research subject you can call Julie Simpson in the UNH Office of Sponsored Research, 603-862-2003, to discuss them.

10. You understand that your age, gender, the characteristics of your speech, and your experience in working with computers or the Project54 system will be recorded, and may be shared with other researchers, along with the data collected about your interactions.

11. You certify that you have read and fully understand the purpose of this research project and the risks and benefits it presents to you as stated above.

I, _____ CONSENT/AGREE to
participate in this research project.

I, _____ REFUSE/DO NOT AGREE
to participate in this research project.

Signature of Subject

Date

I, _____ CONSENT/AGREE to allow
sharing video data with other researchers, including still shots from videos in scientific
publications and technical reports, and showing video data at conferences and similar
meetings.

I, _____ REFUSE/DO NOT AGREE
to allow sharing video data with other researchers or showing it at conferences and
similar meetings.

Signature of Subject

Date

I, _____ CONSENT/AGREE to allow sharing video data with the public by posting video clips, or still shots from the clips, online, or by including them in printed publications.

I, _____ REFUSE/DO NOT AGREE to allow sharing video data with the public by posting video clips, or still shots from the clips, online, or by including them in printed publications.

Signature of Subject

Date

B.5 NASA-TLX Description Presented to Participants

NASA-TLX questionnaire consists of six scales. Table B.3 provides the description of the scales (adapted from [148]) which was handed to participants each time they were required to fill out the NASA-TLX questionnaire.

Scale	Endpoints	Description
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	Low/High	How insecure, discouraged, irritated, stressed, and annoyed or secure, gratified, content, relaxed, and complacent did you feel during the task?

Table B.3 Description of NASA-TLX scales.

APPENDIX C

CHECKING ASSUMPTIONS BEHIND CROSS- CORRELATION MODELS

This appendix provides various graphs which were generated in Chapter 4 for the purpose of analyzing potential problems with the dataset used in modeling our cross-correlation results. Specifically, for each reference study (highway and city driving), cross-correlation result (cumulative and per-glance) and underlying driving performance variables which were used in obtaining the cross-correlation results (steering wheel angle and lane position) we present the following graphs: normal quantile plot, box plot and histogram of studentized residuals as well as residuals versus predicted plots.

Studentized residuals should be distributed as close as possible to a normal distribution. This assumption is satisfied if the data approximately follows a straight line (secondary diagonal line presented in each normal quantile plot) and if the histogram and the box plot are approximately symmetric about 0.

Residuals versus predicted plots are used to check for heteroscedasticity and missing variables. If none of these two problems are present, the data points should be

distributed approximately randomly about the 0 point on the vertical axis (residual axis). If the spread of the data points appears very different in one part of the graph compared to the other, heteroscedasticity may be a problem. On the other hand if the structure of the data points indicates some non-random patterns, it is a sign that the model does not account for all important trends in the data and that more explanatory variables should be included in the model.

As we will see in the graphs presented in the following sections, the distributions of studentized residuals were always very close to normal and no problems have been observed regarding heteroscedasticity and missing variables for any of our regression models.

C.1 Testing Cumulative Steering Wheel Angle Cross-Correlation

Models for Highway Study

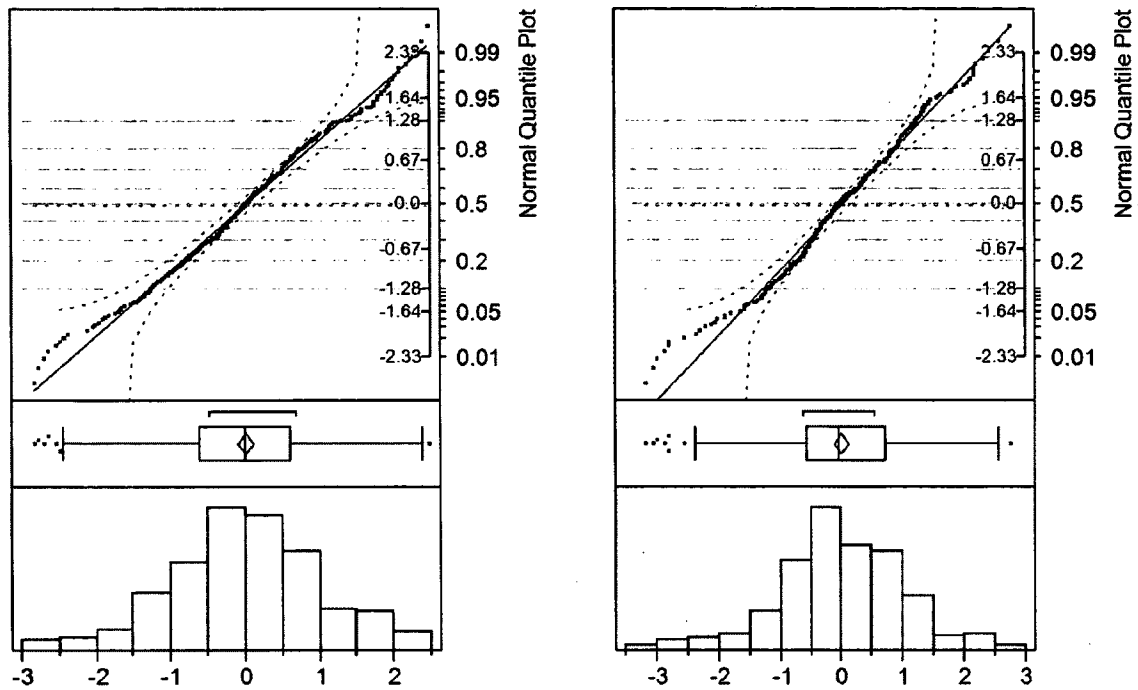


Figure C.29 Highway study: Distributions of studentized residuals for CML_M1_SWA (left) and CML_M2_SWA (right).

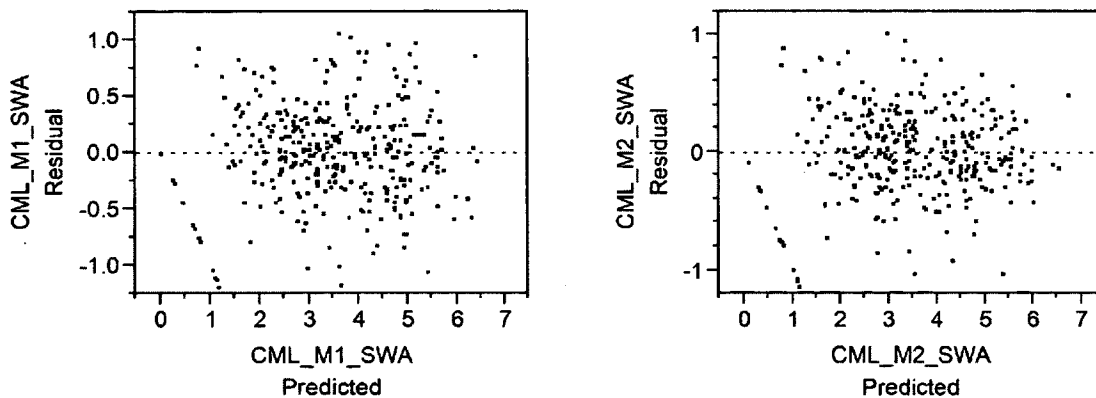


Figure C.30 Highway study: Residuals versus predicted plots for CML_M1_SWA (left) and CML_M2_SWA (right).

C.2 Testing Cumulative Lane Position Cross-Correlation Models **for Highway Study**

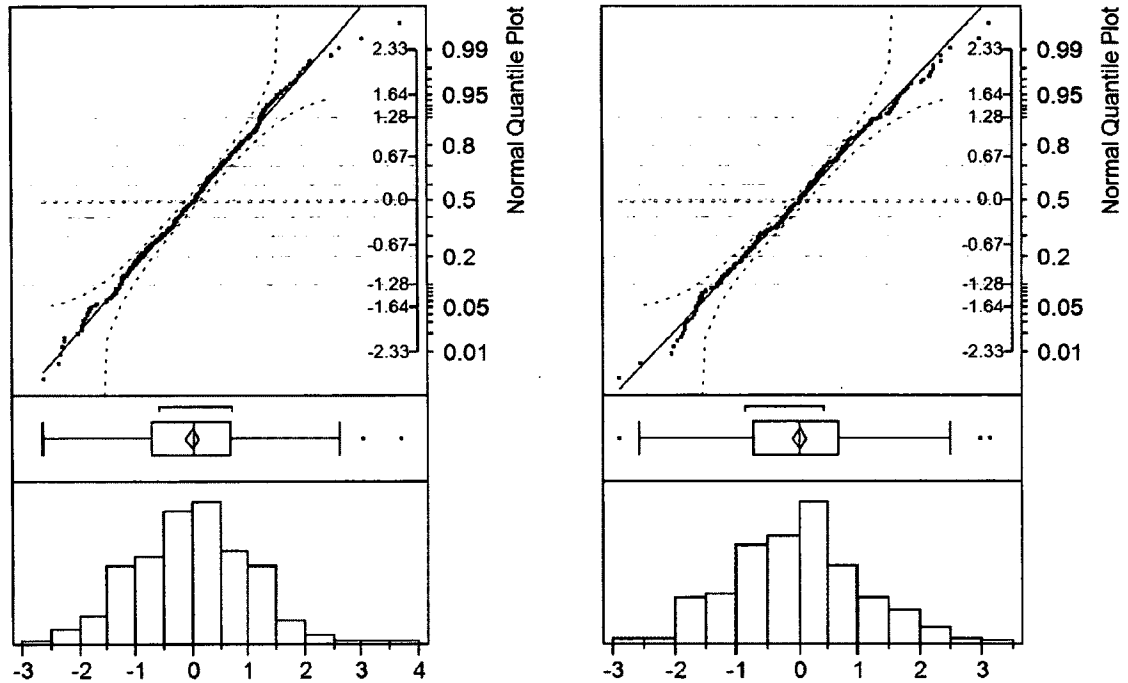


Figure C.31 Highway study: Distributions of studentized residuals for CML_M1_LP (left) and CML_M2_LP (right).

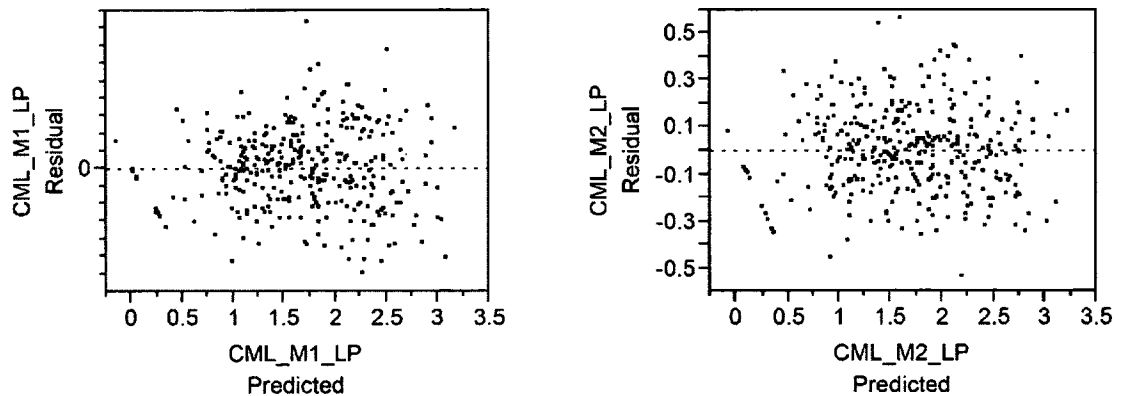


Figure C.32 Highway study: Residuals versus predicted plots for CML_M1_LP (left) and CML_M2_LP (right).

C.3 Testing Cumulative Steering Wheel Angle Cross-Correlation

Models for City Study

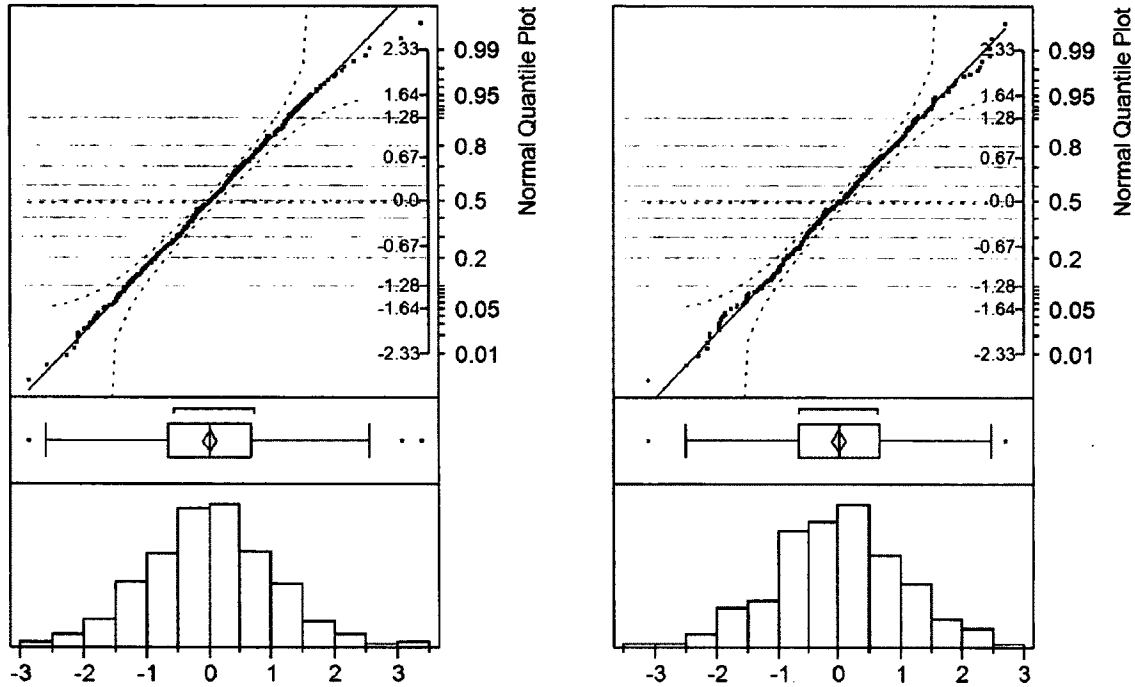


Figure C.33 City study: Distributions of studentized residuals for CML_M1_SWA (left) and CML_M2_SWA (right).

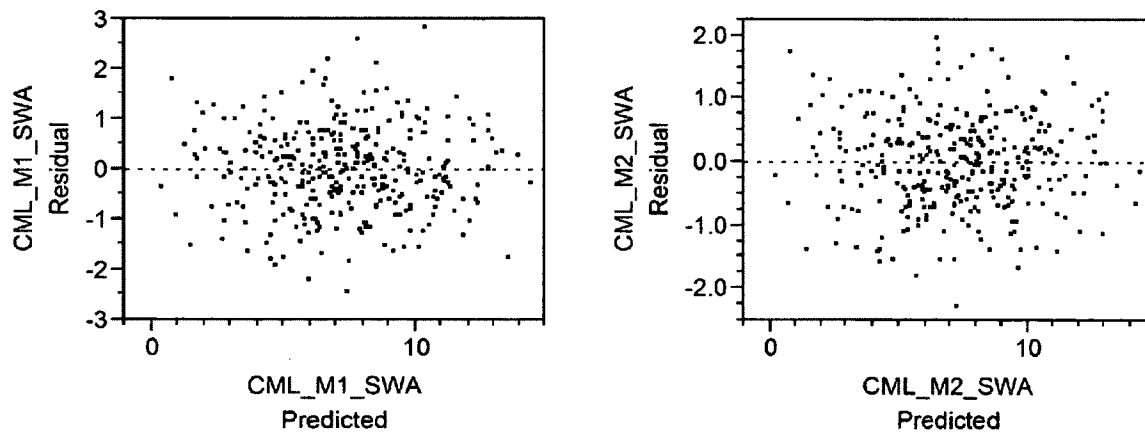


Figure C.34 City study: Residuals versus predicted plots for CML_M1_SWA (left) and CML_M2_SWA (right).

C.4 Testing Cumulative Lane Position Cross-Correlation Models for City Study

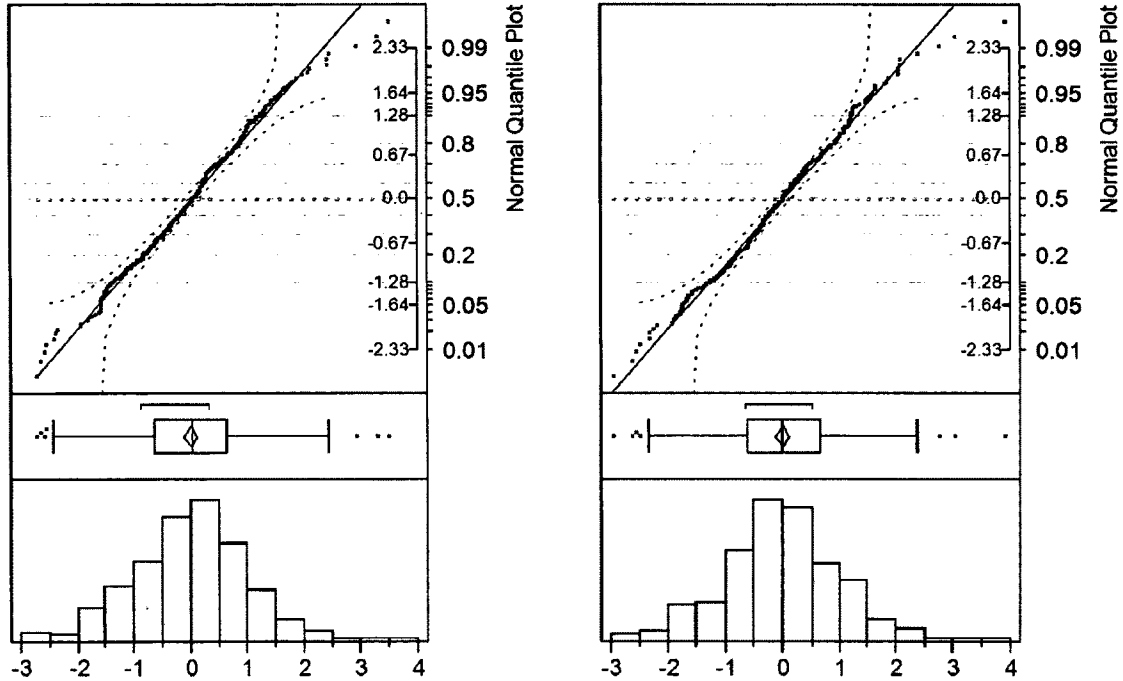


Figure C.35 City study: Distributions of studentized residuals for CML_M1_LP (left) and CML_M2_LP (right).

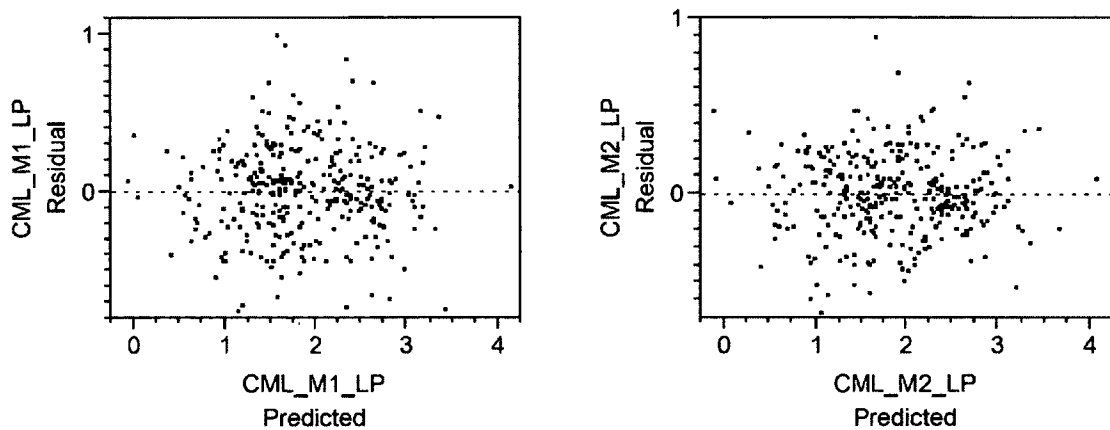


Figure C.36 City study: Residuals versus predicted plots for CML_M1_LP (left) and CML_M2_LP (right).

C.5 Testing Per-Glance Steering Wheel Angle Cross-Correlation

Models for Highway Study

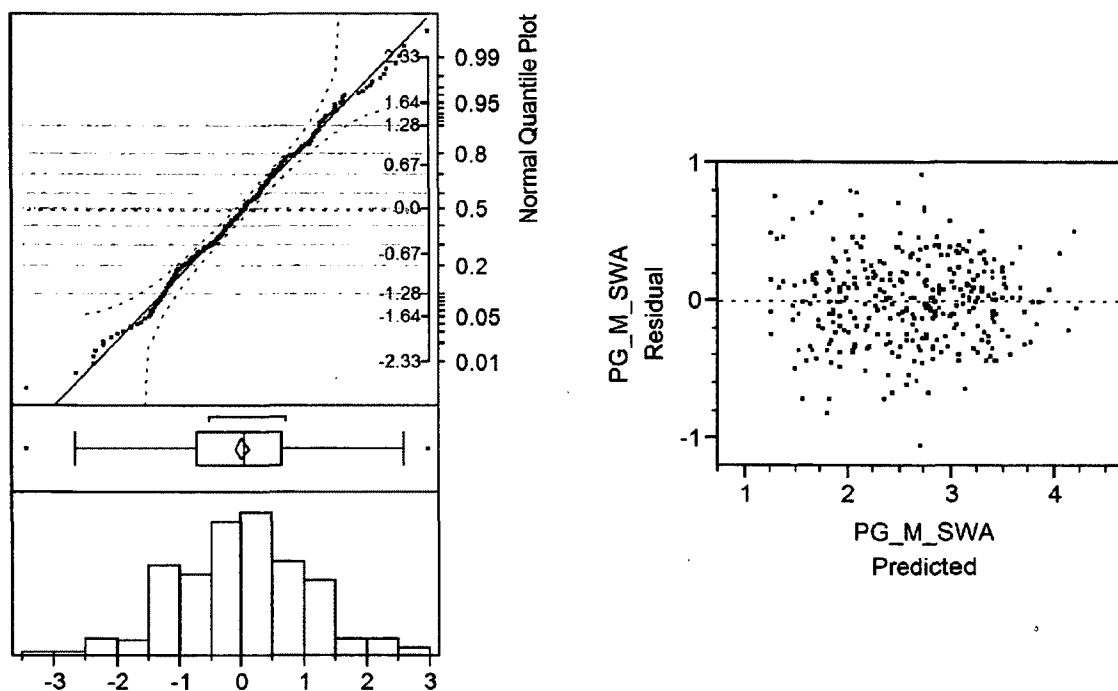


Figure C.37 Highway study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_SWA model.

C.6 Testing Per-Glance Lane Position Cross-Correlation Models

for Highway Study

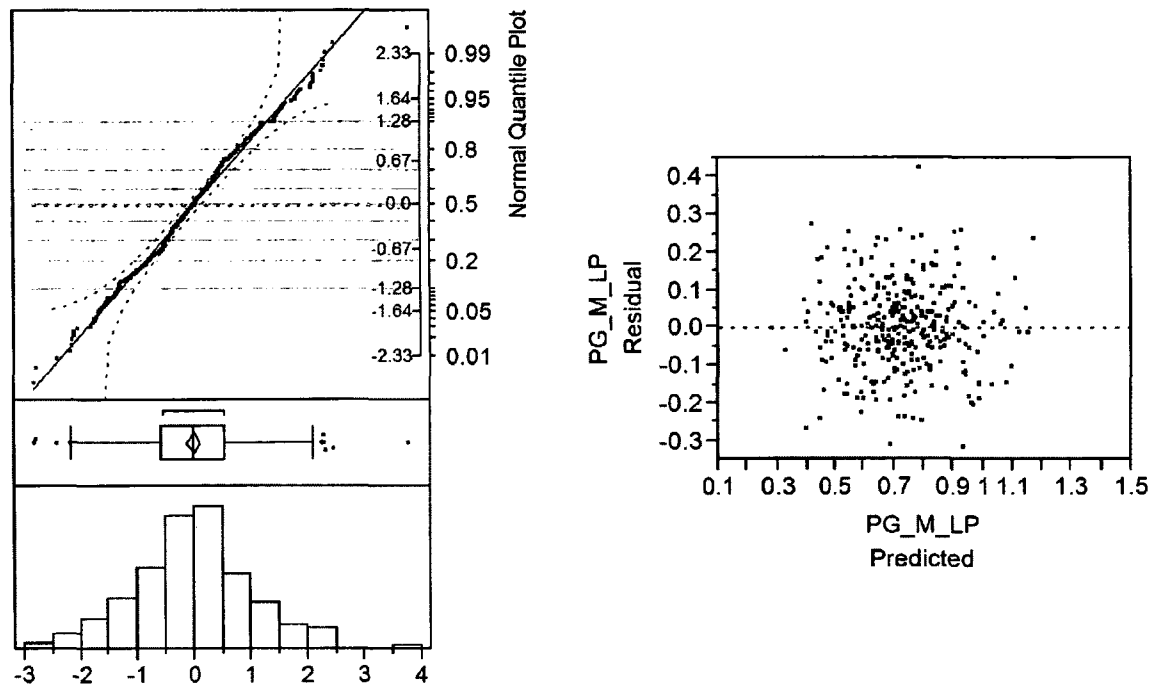


Figure C.38 Highway study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_LP model.

C.7 Testing Per-Glance Steering Wheel Angle Cross-Correlation

Models for City Study

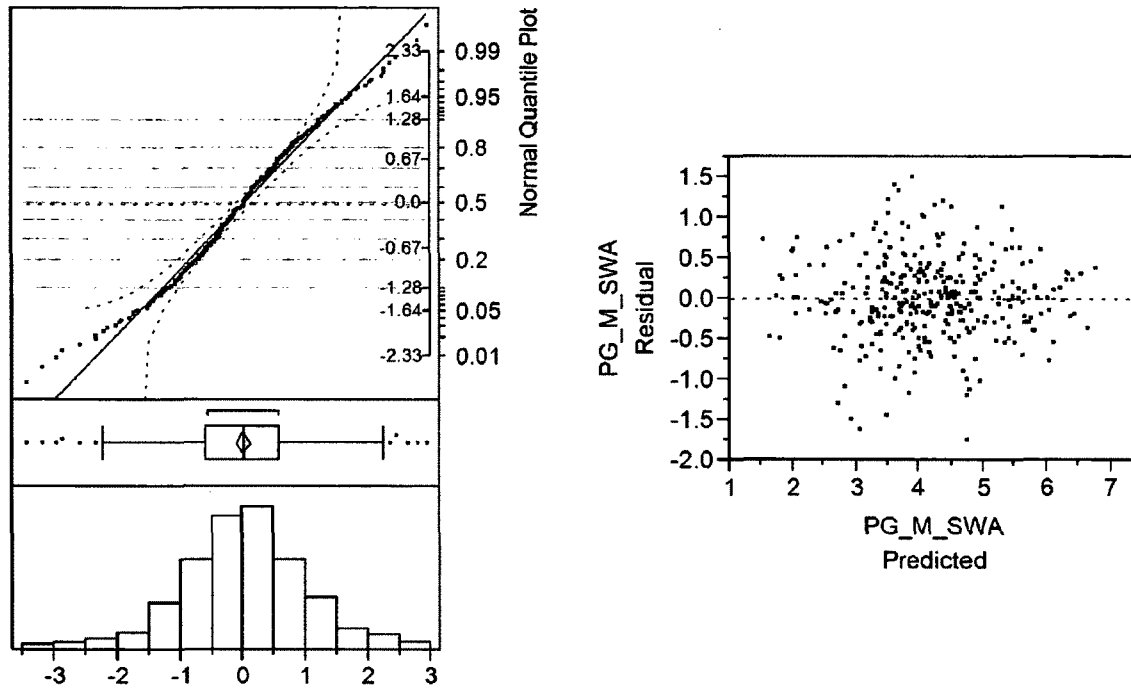


Figure C.39 City study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_SWA model.

C.8 Testing Per-Glance Lane Position Cross-Correlation Models

for City Study

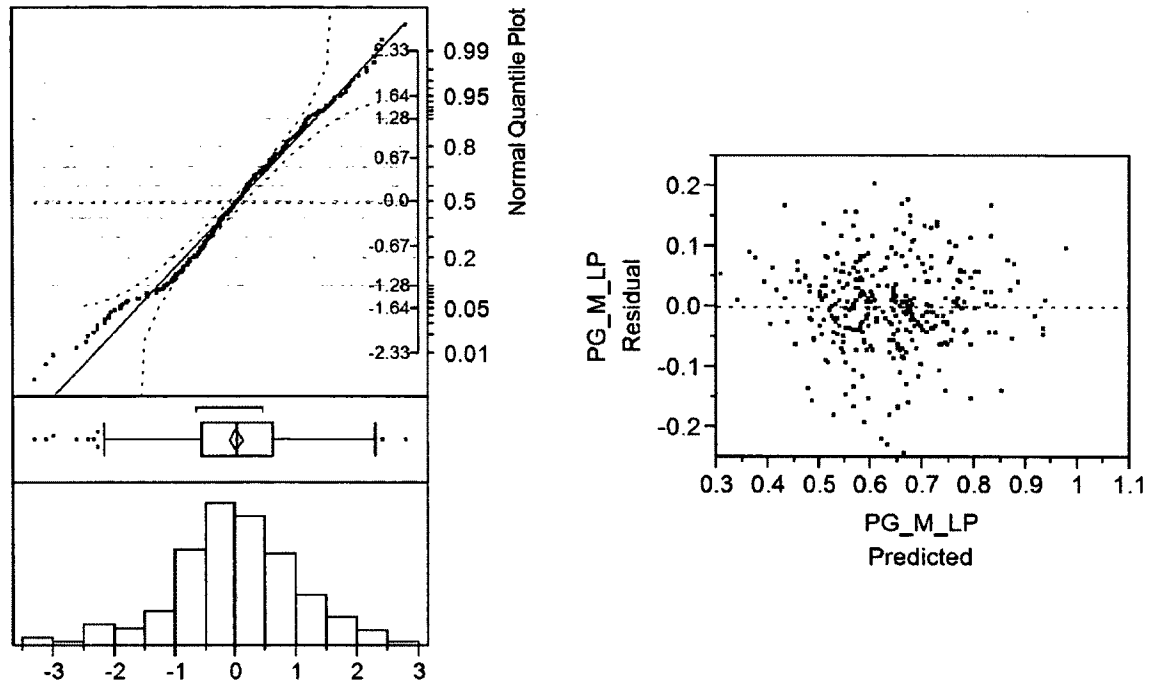


Figure C.40 City study: Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_LP model.

C.9 Testing Cumulative Steering Wheel Angle Cross-Correlation

Models for Pooled Highway and City Studies

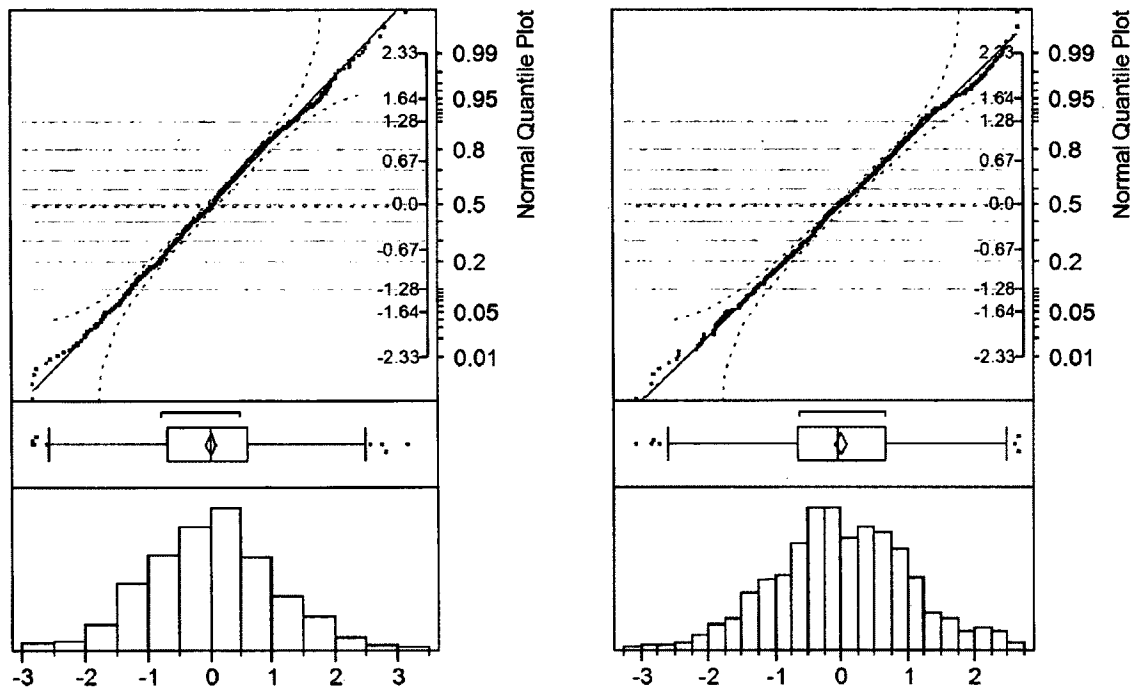


Figure C.41 Distributions of studentized residuals for CML_M1_SWA (left) and CML_M2_SWA (right) models.

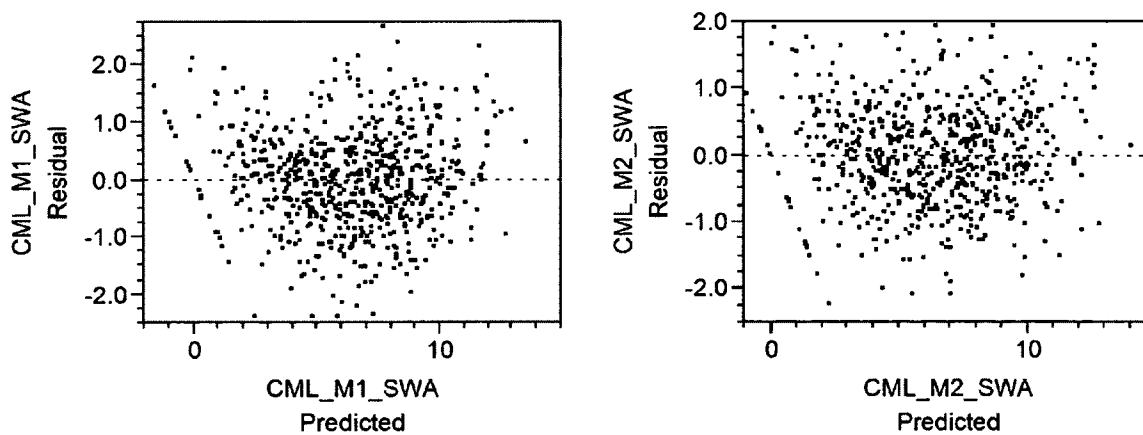


Figure C.42 Residuals versus predicted plots for CML_M1_SWA (left) and CML_M2_SWA (right) models.

C.10 Testing Cumulative Lane Position Cross-Correlation

Models for Pooled Highway and City Studies

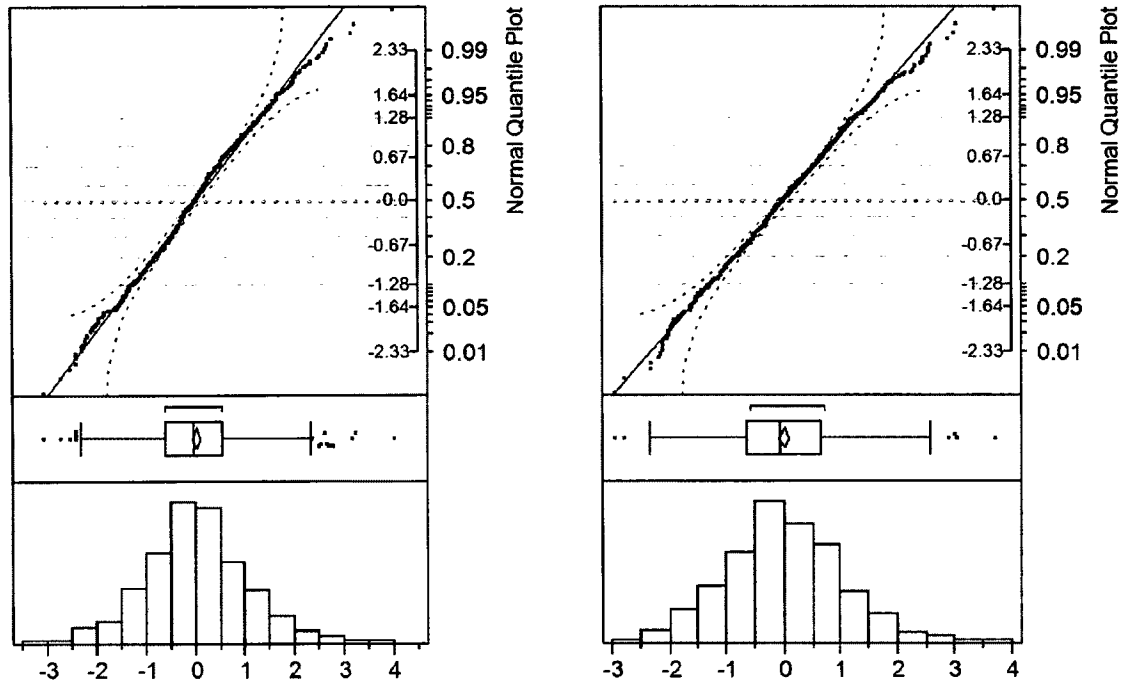


Figure C.43 Distributions of studentized residuals for CML_M1_LP (left) and CML_M2_LP (right) models.

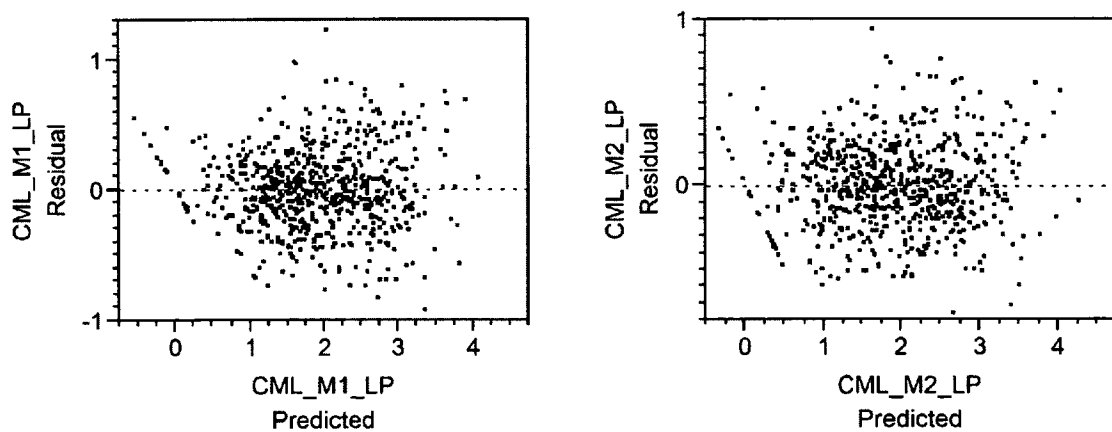


Figure C.44 Residuals versus predicted plots for CML_M1_LP (left) and CML_M2_LP (right).

C.11 Testing Per-Glance Steering Wheel Angle Cross-Correlation

Models for Pooled Highway and City Studies

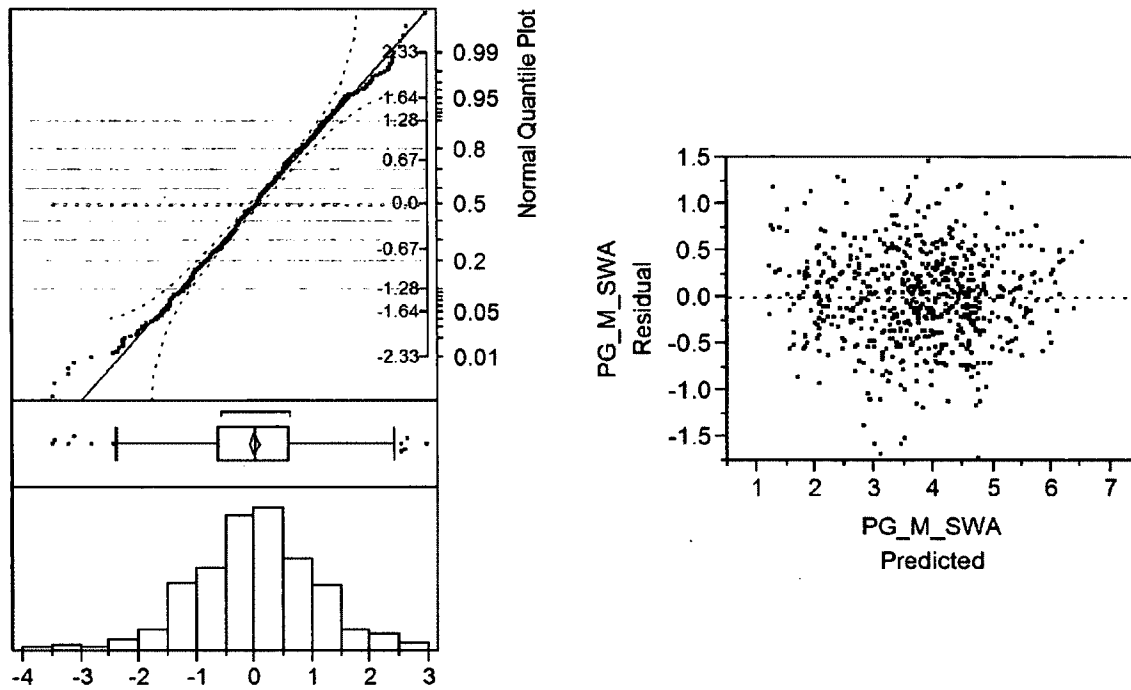


Figure C.45 Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_SWA model.

C.12 Testing Per-Glance Lane Position Cross-Correlation Models **for Pooled Highway and City Studies**

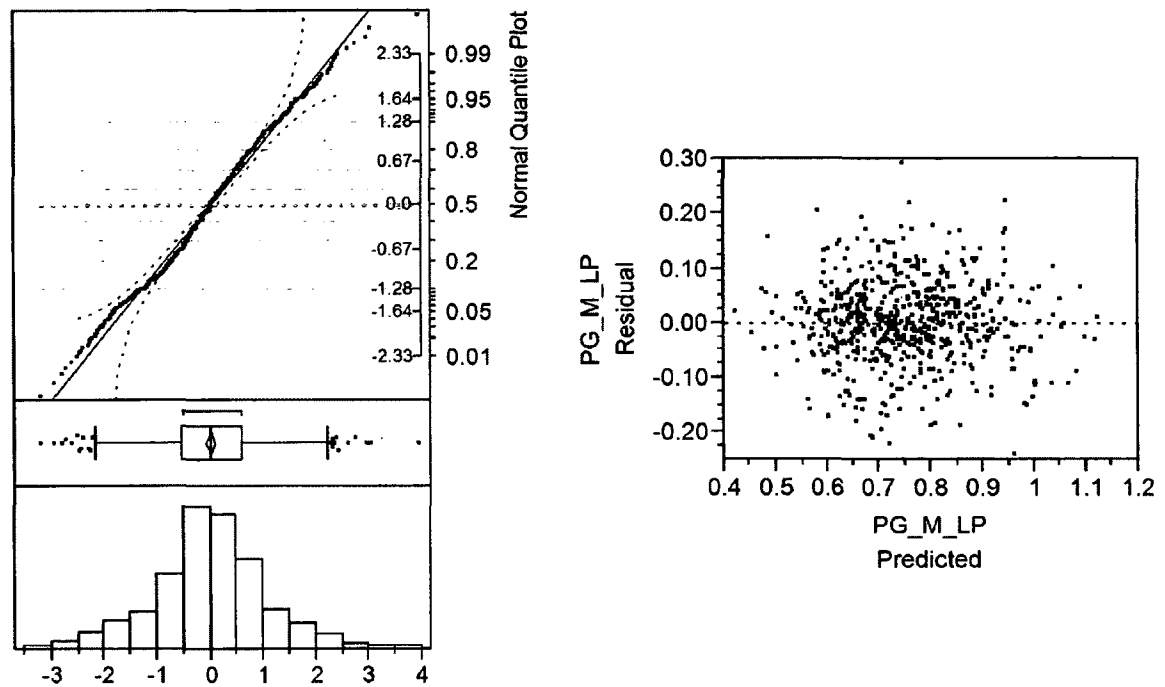


Figure C.46 Distribution of studentized residuals (left) and residuals versus predicted plot (right) for PG_M_LP model.

REFERENCES

- [1] B. McKenzie and M. Rapino, "Commuting in the United States: 2009," U.S. Census Bureau, Washington, DC, American Community Survey Reports, Technical report ACS-15, November 2011.
- [2] G. Leen and D. Heffernan, "Expanding Automotive Electronic Systems," *IEEE Computer*, 35, 1 (2002), 88-93.
- [3] J. C. Stutts, D. W. Reinfurt, L. Staplin, and E. A. Rodgman, "The Role of Driver Distraction in Traffic Crashes," AAA Foundation for Traffic Safety, 2001.
- [4] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, "The Impact of Driver Inattention On Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data," US Department of Transportation, National Highway Traffic Safety Administration (NHTSA), Washington, DC, Technical report DOT HS 810 594, August 2005.
- [5] NHTSA, "Distracted Driving 2009," NHTSA's National Center for Statistics and Analysis, Washington, DC, Technical report DOT HS 811 379, September 2010.
- [6] T. M. Pickrell and T. J. Ye, "Driver Electronic Device Use in 2009," NHTSA's National Center for Statistics and Analysis, Washington, DC, Technical report DOT HS 811 372, September 2010.
- [7] D. Ascone, T. Lindsey, and C. Varghese, "An Examination of Driver Distraction as Recorded in NHTSA Databases," NHTSA's National Center for Statistics and Analysis, Technical report DOT HS 811 216, September 2009.
- [8] G. J. S. Wilde, "Target Risk: Dealing with the danger of death, disease and damage in everyday decisions," 1st edition, Toronto, Ontario, Canada: PDE Publications, 1994.
- [9] D. Gopher and E. Donchin, "Workload - An Examination of the Concept," in *Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance*. K. R. Boff, L. Kaufman, and J. P. Thomas, Eds. John Wiley & Sons, Inc., 1986, pp. 41-1-41-49.
- [10] C. D. Wickens, "Multiple Resources and Performance Prediction," *Theoretical Issues in Ergonomics Science*, 3, 2 (2002), 159-177.

- [11] R. D. Colonel O'Donnell and F. T. Eggemeier, "Workload Assessment Methodology," in *Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance*. K. R. Boff, L. Kaufman, and J. P. Thomas, Eds. John Wiley & Sons, Inc., 1986, 42-1-42-49.
- [12] O. Tsimhoni, D. Smith, and P. Green, "Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard," *Human Factors*, 46, 4 (2004), 600-610.
- [13] R. F. Slick, E. T. Cady, and T. Q. Tran, "Workload Changes in Teenaged Drivers Driving With Distractions," in *Proc. Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (2005), 158-164.
- [14] J. Lai, K. Cheng, P. Green, and O. Tsimhoni, "On the Road and on the Web? Comprehension of synthetic and human speech while driving," in *Proc. CHI 2001* (2001), 206-212.
- [15] W. J. Horrey, C. D. Wickens, and K. P. Consalus, "Modeling Drivers' Visual Attention Allocation while Interacting with In-Vehicle Technologies," *Experimental Psychology: Applied*, 12 (2006), 67-78.
- [16] Y.-C. Liu, "Effects of Using Head-Up Display in Automobile Context on Attention Demand and Driving Performance," *Displays*, 24 (2003), 157-165.
- [17] H. Alm and L. Nilsson, "Changes in Driver Behavior as a Function of Hands-Free Mobile Phones - A Simulator Study," *Accident Analysis & Prevention*, 26, 4 (1994), 441-451.
- [18] D. D. Salvucci, D. Merkle, M. Zuber, and D. P. Brumby, "iPod Distraction: Effects of Portable Music-Player Use on Driver Performance," in *Proc. CHI 2007*, ACM Press (2007), 243-250.
- [19] D. L. Strayer, F. A. Drews, and D. J. Crouch, "A Comparison of the Cell Phone Driver and the Drunk Driver," *Human Factors and Ergonomics Society*, 48, 2 (2006), 381-391.
- [20] J. L. Harbluk, J. S. Mitroi, and P. C. Burns, "Three Navigation Systems with Three Tasks: Using the Lane-Change Test (LCT) to Assess Distraction Demand," in *Proc. Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment and Design* (2009), 24-30.
- [21] J. L. Harbluk, Y. I. Noy, P. L. Trbovich, and M. Eizenman, "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance," *Accident Analysis & Prevention*, 39, 2 (2007), 372-379.
- [22] A. Kun, T. Peak, and Z. Medenica, "The Effect of Speech Interface Accuracy on Driving Performance," in *Proc. Interspeech 2007* (2007).

- [23] A. L. Kun, T. Paek, Z. Medenica, N. Memarovic, and O. Palinko, "Glancing at Personal Navigation Devices Can Affect Driving: Experimental Results and Design Implications," in *Proc. Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ACM (2009), 129-136.
- [24] G. Weinberg, B. Harsham, C. Forlines, and Z. Medenica, "Contextual Push-To-Talk: Shortening Voice Dialogs to Improve Driving Performance," in *Proc. MobileHCI 2010*, ACM Press (2010), 113-122.
- [25] Z. Medenica and A. L. Kun, "Comparing the Influence of Two User Interfaces for Mobile Radios on Driving Performance," in *Proc. Driving Assessment 2007* (2007).
- [26] S. L. Chisholm, J. K. Caird, J. Lockhart, L. Fern, and E. Teteris, "Driving Performance While Engaged In MP-3 Player Interaction: Effects of Practice and Task Difficulty On PRT and Eye Movements," in *Proc. Driving Assessment 2007* (2007).
- [27] D. L. Strayer, J. M. Cooper, and F. A. Drews, "What Do Drivers Fail to See When Conversing on a Cell Phone?" in *Proc. Human Factors and Ergonomics Society 48th Annual Meeting* (2004), 2213-2217.
- [28] S. J. Kim and A. K. Dey, "Simulated Augmented Reality Windshield Display as a Cognitive Mapping Aid for Elder Driver Navigation," in *Proc. CHI 2009* (2009), 133-142.
- [29] M. A. Recarte and L. M. Nunes, "Effects of Verbal and Spatial-Imagery Tasks on Eye Fixations While Driving," *Journal of Experimental Psychology: Applied*, 6, 1 (2000), 31-43.
- [30] H. Zhang, M. R. H. Smith, and G. J. Witt, "Identification of Real-Time Diagnostic Measures of Visual Distraction With an Automatic Eye-Tracking System," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48 (2006), 805-821.
- [31] B. Reimer, B. Mehler, J. F. Coughlin, K. M. Godfrey, and C. Tan, "An On-Road Assessment of the Impact of Cognitive Workload on Physiological Arousal in Young Adult Drivers," in *Proc. 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2009).
- [32] B. Mehler, B. Reimer, J. F. Coughlin, and J. A. Dusek, "The Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers," in *Proc. Transportation Research Board 88th Annual Meeting* (2009).
- [33] R. W. Backs, J. K. Lenneman, J. M. Wetzel, and P. Green, "Cardiac Measures of Driver Workload during Simulated Driving with and without Visual Occlusion,"

Human Factors: The Journal of the Human Factors and Ergonomics Society, 45 (2003), 525-538.

- [34] S. Hosking, K. Young, and M. Regan, "The Effects of Text Messaging on Young Novice Driver Performance," Monash University, Accident Research Centre, Melbourne, Australia, Technical report 246, 2006.
- [35] Y.-C. Liu and M.-H. Wen, "Comparison of head-up display (HUD) vs. head-down display (HDD): driving performance of commercial vehicle operators in Taiwan," *International Journal of Human-Computer Studies*, 61, 5 (2004), 679-697.
- [36] Z. Medenica, A. L. Kun, T. Paek, and O. Palinko, "Augmented Reality vs. Street Views: A Driving Simulator Study Comparing Two Emerging Navigation Aids," in *Proc. MobileHCI 2011*, ACM Press (2011), 265-274.
- [37] M. P. Reed and P. A. Green, "Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialing task," *Ergonomics*, 42, 8 (1999), 1015-1037.
- [38] Kun, Andrew L., Miller, Thomas W. III, and Lenharth, William H, "Computers in Police Cruisers," *IEEE Pervasive Computing*, 3, 4 (2004), 34-41.
- [39] T. A. Ranney, J. L. Harbluk, and Y. Ian Noy, "Effects of Voice Technology on Test Track Driving Performance: Implications for Driver Distraction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47 (2005), 439-454.
- [40] A. L. Kun, T. Paek, Z. Medenica, J. Oppelaar, and O. Palinko, "The Effect of In-Car Navigation Aids on Driving Performance and Visual Attention," Technical report ECE.P54.2009.3, May 2009.
- [41] D. Kern and A. Schmidt, "Design Space for Driver-based Automotive User Interfaces," in *Proc. AutomotiveUI 2009* (2009).
- [42] J. C. Nunnally and I. H. Bernstein, "Psychometric theory," 3rd ed. New York: McGraw-Hill, 1994.
- [43] Lowe, Shelly, "Many Workers Have Long Commutes to Work," US Census Bureau Press Release. Date accessed 5/30/2009.
- [44] M. Bruce and B. Deborah, "Improved Vehicle Safety and How Technology Will Get Us There, Hopefully," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*. K. Takeda, H. Erdogan, J. H. L. Hansen, and H. Abut, Eds. Springer US, 2009, 1-8.
- [45] J. D. Lee and D. L. Strayer, "Preface to the Special Section on Driver Distraction," *Human Factors*, 46, 4 (2004), 583-586.

- [46] J. A. Michon, "A critical view of driver behavior models: What do we know, what should we do?" in *Human behavior and traffic safety*. L.Evans and R.C.Schwing, Eds. New York: Plenum, 1985, 485-520.
- [47] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman, "An Overview of the 100-Car Naturalistic Study and Findings," Technical report 05-0400, 2009.
- [48] T. B. Sheridan, "Driver Distraction From a Control Theory Perspective," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46 (2004), 587-599.
- [49] K. Young, M. Regan, and M. Hammer, "Driver Distraction: A Review of the Literature," Monash University, Accident Research Centre, Melbourne, Australia, Technical report 206, 2003.
- [50] B. Zylstra, O. Tsimhoni, P. Green, and K. Mayer, "Driving Performance for Dialing, Radio Tuning, and Destination Entry while Driving Straight Roads," The University of Michigan Transportation Research Institute, Ann Arbor, MI, Technical report UMTRI-2003-35, 2003.
- [51] M. Langham, G. Hole, J. Edwards, and C. O'Neil, "An analysis of 'looked but failed to see' accidents involving parked police vehicles," *Ergonomics*, 45, 3 (2002), 167-185.
- [52] B. Reimer, "Cognitive Task Complexity and the Impact on Drivers' Visual Tunneling," in *Proc. Transportation Research Board 88th Annual Meeting* (2008).
- [53] B. Reimer, B. Mehler, J. F. Coughlin, N. Roy, and J. A. Dusek, "The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups," *Transportation Research Part F*, 14, 1 (2010), 13-25.
- [54] S. Amado and P. Ulupinar, "The Effects of Conversation on Attention and Peripheral Detection: Is Talking With a Passenger and Talking on the Cell Phone Different?" *Transportation. Research. Part F. : Traffic. Psychology. and Behaviour*, 8, 6 (2005), 383-395.
- [55] D. L. Strayer and F. A. Drews, "Profiles in Driver Distraction: Effects of Cell Phone Conversations on Younger and Older Drivers," *Human Factors*, 46, 4 (2004), 640-649.
- [56] B. Wallace, "Driver Distraction by Advertising: Genuine Risk or Urban Myth?" *Municipal Engineer*, 156, ME3 (2003), 185-190.
- [57] K. Rumar, "Impacts on road design of the human factor and information systems," in *Proc. 9th International Road Federation Meeting* (1981), 31-49.

- [58] R. M. Yerkes and J. D. Dodson, "The Relation of Strength of Stimulus to Rapidity of Habit-Formation," *Journal of Comparative Neurology and Psychology*, 18 (1908), 459-482.
- [59] F. Nachreiner, "Standards for ergonomics principles relating to the design of work systems and to mental workload," *Applied Ergonomics*, 26, 4 (1995), 259-263.
- [60] D. L. Strayer and F. A. Drews, "Multi-tasking in the automobile," in A. Kramer, D. Wiegmann, & A. Kirlik (Eds.), *Applied attention: From theory to practice* (2006), 121-133.
- [61] C. D. Wickens, "Multiple Resources and Mental Workload," *Human Factors*, 50, 3 (2008), 449-455.
- [62] D. Navon and D. Gopher, "On the economy of the human-processing system," *Psychological Review*, 86, 3 (1979), 214-255.
- [63] D. Broadbent, "Perception and Communications," New York: Pergamon Press, 1958.
- [64] A. T. Welford, "Single Channel Operation in the Brain," *Acta Psychologica*, 27 (1967), 5-21.
- [65] D. Kahneman, "Attention and Effort," Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- [66] J. T. Reason, "Managing the Risks of Organizational Accidents," 1st ed., Ashgate Publishing Limited, 1997.
- [67] W. W. Wierwille and L. Tijerina, "An Analysis of Driving Accident Narratives As a Means of Determining Problems Caused by In-Vehicle Visual Allocation and Visual Workload," in Gale, A.G., Brown, I.D., Haslegrave, C.M., and Taylor, S.P. (Eds.), *Vision in Vehicles V*, Amsterdam, The Netherlands: Elsevier, 1996, 79-86.
- [68] P. Green, "Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions," in *Proc. of the 2004 International Congress on Transportation Electronics (Convergence 2004, SAE publication P-387)*, Society of Automotive Engineers (2004), 165-180.
- [69] B. Mehler, B. Reimer, and J. F. Coughlin, "Physiological Reactivity to Graded Levels of Cognitive Workload across Three Age Groups: An On-Road Evaluation," in *Proc. of the Human Factors and Ergonomics Society 54th Annual Meeting* (2010), 2062-2066.
- [70] F. T. Eggemeier and G. F. Wilson, "Performance-based and subjective assessment of workload in multi-task environments," in *Multiple-task performance*. D. L. Damos, Ed. London: Taylor & Francis, 1991, 217-278.

- [71] H. Alm and L. Nilsson, "The Effects of a Mobile Telephone Task on Driver Behaviour in a Car Following Situation," *Accident. Analysis. & Prevention.*, 27, 5 (1995), 707-715.
- [72] W. J. Horrey, C. D. Wickens, and A. L. Alexander, "The Effects of Head-Up Display Clutter and In-Vehicle Display Separation on Concurrent Driving Performance," in *Proc. of the Human Factors and Ergonomics Society 47th Annual Meeting* (2003).
- [73] D. L. Strayer and W. A. Johnston, "Driven to Distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone," *Psychological Science*, 12, 6 (2001), 462-466.
- [74] K. F. Van Orden, W. Limbert, S. Makeig, and T. Jung, "Eye Activity Correlates of Workload during a Visuospatial Memory Task," *Human Factors*, 43, 1 (2001), 111-121.
- [75] O. Palinko and A. L. Kun, "Exploring the Influence of Light and Cognitive Load on Pupil Diameter in Driving Simulator Studies," in *Proc. Driving Assessment 2011* (2011).
- [76] A. Riener, A. Ferscha, and M. Aly, "Heart on the Road: HRV Analysis for Monitoring a Driver's Affective State," in *Proc. AutomotiveUI 2010* (2010).
- [77] R. McCraty, M. Atkinson, W. Tiller, G. Rein, and A. D. Watkins, "The Effects of Emotions on Short Term Power Spectrum Analysis of Heart Rate Variability," *The American Journal of Cardiology*, 76, 14 (1995), 1089-1093.
- [78] G. F. Wilson, "An analysis of mental workload in pilots during flight using multiple psychophysiologic measures," *The International Journal of Aviation Psychology*, 12, 1 (2002), 3-18.
- [79] W. Boucsein, "Electrodermal Activity," 2nd ed. New York, NY 10013, USA: Springer Science + Business Media, LLC, 2012.
- [80] Cacioppo, John T. and Tassinari, Louis G, "Inferring Psychological Significance From Physiological Signals," *American Physiologist*, 45, 1 (1990), 16-28.
- [81] P. Trbovich and J. L. Harbluk, "Cell phone communication and driver visual behavior: the impact of cognitive distraction," *ACM Press* (2003), 728-729.
- [82] L. M. Nunes and M. A. Recarte, "Cognitive demands of hands-free-phone conversation while driving," *Transportation. Research. Part F. : Traffic. Psychology. and Behaviour*, 5, 2 (2002), 133-144.
- [83] Society of Automotive Engineers, "SAE J2396 Recommended Practice," July 2000.

- [84] G. E. Cooper and R. P. Jr. Harper, "The use of pilot rating in the evaluation of aircraft handling qualities," Moffett Field, CA: Ames Research Center, National Aeronautics and Space Administration, Technical report TN-D-5153, 1969.
- [85] G. B. Reid, F. T. Eggemeier, and C. A. Shingledecker, "Subjective workload assessment technique," in *Proc. AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics* (1982), 281-288.
- [86] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*. P. A. Hancock and N. Meshkati, Eds. North Holland Press, 1988, 239-250.
- [87] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Drivers' Attention to the Roadway," *Human Factors*, 43 (2001), 631-640.
- [88] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs, and J. Brown, "Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance," *Accident Analysis & Prevention*, 38, 1 (2006), 185-191.
- [89] T. C. Lansdown, N. Brook-Carter, and T. Kersloot, "Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications," *Ergonomics*, 47, 1 (2004), 91-104.
- [90] A. Pauzié, "A method to assess the driver mental workload: The driving activity load index (DALI)," *IET Intelligent Transportation Systems*, 2, 4 (2008), 315-322.
- [91] D. Kern, A. Mahr, S. Castronovo, A. Schmidt, and C. Müller, "Making Use of Drivers' Glances onto the Screen for Explicit Gaze-Based Interaction," in *Proc. AutomotiveUI 2010* (2010), 110-113.
- [92] E. Chin, F. Nathan, A. Pauzié, J. Manzano, E. Nodari, C. Cherri, A. Rambaldini, A. Toffetti, and M. Marchitto, "Subjective Assessment Methods for Workload," Technical report AIDE IST-1-507674-IP, 2006.
- [93] S. Jamieson, "Likert scales: how to (ab)use them," *Medical Education* 2004, 38 (2004), 1212-1218.
- [94] R. Bader, O. Siegmund, and W. Woerndl, "A Study on User Acceptance of Proactive In-Vehicle Recommender Systems," in *Proc. AutomotiveUI 2011* (2011), 47-54.
- [95] P. Fröhlich, M. Baldauf, M. Hagen, S. Suetterle, D. Schabus, and A. L. Kun, "Investigating Safety Services on the Motorway: the Role of Realistic Visualization," in *Proc. AutomotiveUI 2011* (2011), 143-150.

- [96] S. Consolvo and M. Walker, "Using the Experience Sampling Method to Evaluate Ubicomp Applications," *IEEE Pervasive Computing*, 2, 2 (2003), 24-31.
- [97] J. O. Keith and Gary E. Burnett, "Learning-oriented vehicle navigation systems: a preliminary investigation in a driving simulator," *ACM* (2008), 119-126.
- [98] K. Itoh, Y. Miki, N. Yoshitsugu, and N. Kubo, "Evaluation of a Voice-Activated System Using a Driving Simulator," in *Proc. SAE 2004 World Congress & Exhibition* (2004).
- [99] V. Charissis, S. Papanastasiou, and G. Vlachos, "Interface Development for Early Notification Warning System: Full Windshield Head-Up Display Case Study," in *Proc. of the 13th International Conference on Human-Computer Interaction* (2009), 683-692.
- [100] O. Tsimhoni and P. Green, "Visual Demand of Driving Curves as Determined by Visual Occlusion," *Vision in Vehicles*, 8 (1999).
- [101] G. E. Burnett, "A Road-Based Evaluation of a Head-Up Display for Presenting Navigation Information," in *Proc. of HCI International conference* (2003), 180-184.
- [102] M. E. Lesch and P. A. Hancock, "Driving Performance During Concurrent Cell-Phone Use: Are Drivers Aware of Their Performance Decrements," *Journal of Accident Analysis & Prevention*, 36 (2003), 471-480.
- [103] J. Healey and R. W. Picard, "Detecting Stress During Real World Driving Tasks Using Physiological Sensors," *IEEE Transactions on Intelligent Transportation Systems*, 6, 2 (2005), 156-166.
- [104] K. M. Bach, M. G. Jæger, M. B. Skov, and N. G. Thomassen, "Evaluating Driver Attention and Driving Behavior: Comparing Controlled Driving and Simulated Driving," in *Proc. of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction* (2008), 193-201.
- [105] M. Wooldridge, K. Bauer, P. Green, and K. Fitzpatrick, "Comparison of Driver Visual Demand in Test Track, Simulator, and On-Road Environments," in *Proc. Transportation Research Board, 79th Annual Meeting* (2000).
- [106] R. F. Slick, E. Kim, D. F. Evans, and J. P. Steele, "Using Simulators to Train Novice Teen Drivers: Assessing Psychological Fidelity as a Precursor of Transfer of Training," in *Proc. Asian Conference on Driving Simulation 2006* (2006).
- [107] Y. Wang, B. Mehler, B. Reimer, V. Lammers, L. A. D'Ambrosio, and J. F. Coughlin, "The Validity of Driving Simulation for Assessing Differences Between In-Vehicle Informational Interfaces: A Comparison with Field Testing," *Ergonomics*, 53, 3 (2010), 404-420.

- [108] H. L. Lew, J. H. Poole, E. H. Lee, D. L. Jaffe, H. Huang, and E. Brodd, "Predictive Validity of Driving-Simulator Assessments Following Traumatic Brain Injury: A Preliminary Study," *Brain Injury*, 19, 3 (2005), 177-188.
- [109] A. Kemeny and F. Panerai, "Evaluating perception in driving simulation experiments," *Trends Cogn Sci.*, 7, 1 (2003), 31-37.
- [110] Drive Safety, "DS-600c Research Simulator," Drive Safety, 2011.
- [111] M. J. Kelly, S. Lassacher, and Z. Shipstead, "A High Fidelity Driving Simulator as a Tool for Design and Evaluation of Highway Infrastructure Upgrades," Montana Department of Transportation, Technical report FHWA/MT-07-005/8117-33, 2007.
- [112] J. O. Brooks, R. R. Goodenough, M. C. Crisler, N. D. Klein, R. L. Alley, B. L. Koon, J. Logan, J. H. Ogle, R. A. Tyrrell, and R. F. Wills, "Simulator sickness during driving simulation studies," *Accident Analysis & Prevention*, 42, 3 (2010), 788-796.
- [113] K. Torkkola, M. Massey, and C. Wood, "Detecting driver inattention in the absence of driver monitoring sensors," in *Proc. of the 2004 International Conference on Machine Learning and Applications* (2004), 220-226.
- [114] R. Maurant and T. Thattacherry, "Simulator Sickness in a Virtual Environments Driving Simulator," in *Proc. Human Factors and Ergonomics Society, Annual Meeting* (2000), 534-537.
- [115] G. E. Burnett, A. Irune, and A. Mowforth, "Driving simulator sickness and validity: how important is it to use real car cabins?" *Advances in Transportation Studies an international Journal* (2007).
- [116] D. D. Salvucci and K. L. Macuga, "Predicting the Effects of Cell-Phone Dialing on Driver Performance," in *Proc. Fourth International Conference on Cognitive Modeling* (2001).
- [117] R. R. Veit, P. Pyle, and J. A. McGowan, "Ocean Warming and Long-Term Change in Pelagic Bird Abundance within the California Current System," *Marine Ecology Progress Series*, 139 (1996), 11-18.
- [118] D. M. Simpson, A. F. C. Infantosi, and D. A. Botero Rosas, "Estimation and significance testing of cross-correlation between cerebral blood flow velocity and background electro-encephalograph activity in signals with missing samples," *Medical & Biological Engineering & Computing*, 39, 4 (2001), 428-433.
- [119] J. Chen, J. Benesty, and Y. A. Huang, "Time Delay Estimation Using Spatial Correlation Techniques," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC2003)* (2003).

- [120] R. M. Reich, R. L. Czaplewski, and W. A. Bechtold, "Spatial cross-correlation of undisturbed, natural shortleaf pine stands in northern Georgia," *Environmental and Ecological Statistics*, 1 (1994), 201-217.
- [121] J. N. Sarvaiya, S. Patnaiak, and S. Bombaywala, "Image Registration by Template Matching Using Normalized Cross-Correlation," in *Proc. ACT '09* (2009), 819-822.
- [122] O. Tsimhoni and P. Green, "Visual Demand of Driving and the Execution of Display-Intensive In-Vehicle Tasks," in *Proc. of the Human Factors and Ergonomics Society 45th Annual Meeting* (2001), 1586-1590.
- [123] P. Green and K. George, "When Should Auditory Guidance Systems Tell Drivers to Turn," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 39 (1995), 1072-1076.
- [124] B. Reimer, B. Mehler, V. Lammers, Y. Wang, and J. F. Coughlin, "A Comparison of Three Manual Destination Entry Methods on Their Impact on Visual Attention and Driving Performance: An On-Road Evaluation," in *Proc. of the Human Factors and Ergonomics Society 53rd annual meeting* (2009), 1757-1761.
- [125] J. R. Sayer, J. M. Devonshire, and C. A. Flannagan, "The Effects of Secondary Tasks on Naturalistic Driving Performance," Technical report UMTRI-2005-29, 2005.
- [126] S. J. Orfanidis, "Optimum Signal Processing," 2nd ed. New York, NY: McGraw-Hill Publishing Company, 2007.
- [127] Virtual CableTM. <http://www.mvs.net/>. Date accessed 4/10/2012.
- [128] Google Street View. <http://maps.google.com/>. Date accessed 4/10/2012.
- [129] T. A. Dingus, R. J. Hanowski, and S. G. Klauer, "Estimating Crash Risk," *Ergonomics in Design: The Quarterly of Human Factors Applications*, 19, 8 (2011), 8-12.
- [130] Kosinski, Robert J., "A Literature Review on Reaction Time," Online article. Date accessed 09/10/2010.
- [131] M. Green, "How Long Does It Take to Stop? Methodological Analysis of Driver Perception-Brake Times," *Transportation Human Factors*, 2, 3 (2000), 195-216.
- [132] L. C. Hamilton, "Regression with graphics: a second course in applied statistics," Belmont, CA, USA: Wadsworth, Inc., 1992.
- [133] S. Franklin and S. Thomas, "Robust Multivariate Outlier Detection Using Mahalanobis' Distance and Modified Stahel-Donoho Estimators," in *Proc. of The Second International Conference on Establishment Surveys* (2000).

- [134] C. A. Pickering, K. J. Burnham, and M. J. Richardson, "A Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction," in *Proc. Automotive Electronics, 2007 3rd Institution of Engineering and Technology Conference on* (2007), 1-15.
- [135] Ernst, Kurt, "2013 Ford Fusion To Offer Lane Keeping System," MotorAuthority. Date accessed 8/1/2012.
- [136] Mone, Gregory, "Adaptive Cruise Control Goes Mainstream," Wired Magazine. Date accessed 8/1/2012.
- [137] O. Tsimhoni, H. Yoo, and P. Green, "Effects of Visual Demand and In-Vehicle Task Complexity on Driving and Task Performance as Assessed by Visual Occlusion," The University of Michigan Transportation Research Institute, Ann Arbor, MI, Technical report UMTRI-99-37, December 1999.
- [138] A. L. Kun, Z. Medenica, O. Palinko, and P. A. Heeman, "Utilizing Pupil Diameter to Estimate Cognitive Load Changes During Human Dialogue: A Preliminary Study," in *Proc. Cognitive Load and In-Vehicle Human-Machine Interaction workshop 2011* (2011).
- [139] S. T. Iqbal, X. S. Zheng, and B. P. Bailey, "Task-Evoked Pupillary Response to Mental Workload in Human-Computer Interaction," in *Proc. CHI 2004* (2004), 1477-1480.
- [140] Pritchett, Ginnie, "AAA Announces New Vehicle Technology Trends for 2012," AAA, 2012.
- [141] O. Palinko and A. L. Kun, "Prototype Wireless Push-to-Talk Glove," in *Proc. IET 2008* (2008).
- [142] M. Van Der Hulst, T. Meijman, and T. Rothengatter, "Maintaining task set under fatigue: a study of time-on-task effects in simulated driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, 4, 2 (2001), 103-118.
- [143] W. W. Wierwille, S. S. Wreggit, C. L. Kirn, L. A. Ellsworth, and R. J. Fairbanks, "Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness," National Highway Traffic Safety Administration, Technical report DOT HS 808 247, 1994.
- [144] Seeing Machines, "Guide to using NTP FastTrack," November 2005.
- [145] T. W. Miller, III, "Overview of Project54 inter-application messaging: Local and distributed messaging architectures," Technical report ECE.P54.2004.11, September 2004.
- [146] Seeing Machines, "faceLAB 5," Seeing Machines, 2009.

- [147] Thought Technology, "Pro-Comp Infinity," Thought Technology.
- [148] A. Cao, K. K. Chintamani, A. K. Pandya, and R. D. Ellis, "NASA TLX: Software for assessing subjective mental workload," *Behavior Research Methods*, 41, 1 (2009), 113-117.